DOCUMENT RESUME

ED 292 030 CG 020 637

AUTHOR	Shock, Nathan W.; And Others
TITLE	Normal Human Aging: The Baltimore Longitudinal Study of Aging.
INSTITUTION	National Inst. on Aging (DHHS/NIH), Bethesda, MD.
REPORT NO	NIH-84-2450
PUB DATE	Nov 84
NOTE	661p.
AVAILABLE FROM	Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402.
PUB TYPE	Reports - General (140) -- Reports - Research/Technical (143) -- Books (010)
EDRS PRICE	MF03/PC27 Plus Postage.
DESCRIPTORS	*Age Differences; *Aging (Individuals); *Individual Development; *Longitudinal Studies; Research Methodology
IDENTIFIERS	*Baltimore Longitudinal Study of Aging

ABSTRACT

The Baltimore Longitudinal Study of Aging (BLSA) was begun in 1958 to trace the effects of aging in humans and to distinguish between the true effects of aging and those processes, including disease, socioeconomic disadvantage, and lack of educational opportunity, that may appear or become more pronounced with time but are biologically irrelevant to the underlying mechanisms of human aging. It has sought to achieve this goal by making measurements on over 1,000 male volunteers who at their entry into the study were aged 17 to 96 years and were living independently in the community. Since 1978, women have been systematically added to the BLSA. This document presents a report of the BLSA's first 23 years. Chapter I describes methods for the study of aging. Chapter II provides a literature review of studies that illustrate the strengths and weaknesses of the longitudinal method and that were conducted on normal community-residing adults. Chapter III describes the general design of the study, characteristics of the subjects, the tests used, data storage and retrieval, and methods of data analysis. Chapter IV describes in detail the testing procedures followed. The results of published BLSA cross-sectional analyses are summarized in chapter V, while those based on longitudinal examinations of selected variables are summarized in chapter VI, and 35 of those longitudinal studies themselves are reprinted in the appendix, which comprises over half of this volume. Preceding the appendix, the seventh and last chapter looks at future directions for the BLSA. Thirty-five tables and 48 figures are included, and a 386-item reference list is provided. (NB)

****
*   Reproductions supplied by EDRS are the best that can be made   *
*   from the original document.   *
****

NORMAL HUMAN AGING:

THE BALTIMORE LONGITUDINAL STUDY OF AGING

ED292030

CG 020637

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

U.S.
DEPARTMENT
OF
HEALTH
AND
HUMAN
SERVICES

Public
Health
Service

National
Institutes
of
Health

BEST COPY AVAILABLE

Normal Human Aging:

The Baltimore Longitudinal Study of Aging

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

NORMAL HUMAN AGING:

The Baltimore Longitudinal Study of Aging

NATHAN W. SHOCK
Richard C. Greulich
Reubin Andres
David Arenberg
Paul T. Costa, Jr.
Edward G. Lakatta
Jordan D. Tobin

U. S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health
National Institute on Aging
Gerontology Research Center
Baltimore City Hospitals

NIH Publication No. 84-2450
November, 1984

Dedication

Dr. W.W. Peter

This volume is dedicated to Dr. W.W. Peter (Fig. 1), whose foresight and enthusiasm played an important role in the recruitment of subjects for the Baltimore Longitudinal Study of Aging. Dr. Peter, a retired Medical Officer of the United States Public Health Service, provided the initial impetus that led to the recruitment of normal men residing in the community as subjects for their entire lifetimes in a study of aging. In 1958 very few people recognized the impact that the increasing number of elderly people in the population would have on our society. Only a few voices were raised to point out the need for research on aging. Dr. Peter's was one of them.

Perhaps because of his experience as a physician, Dr. Peter saw the need to study aging in individuals who were free from disabling diseases and were leading successful lives in the community. He "sold" this idea to his friends, neighbors, and colleagues and persuaded them to enroll as participants in a study that would involve repeated tests and examinations over their entire life spans. He represented participation in the study as a contribution to science and the future of mankind.

Dr. Peter soon realized that his own efforts were not likely to recruit enough subjects to answer many of the questions posed about aging. Hence he proposed to participants already in the study that they join him in recruiting others among their colleagues and friends. Thanks to Dr. Peter's dedication, the system worked! The waiting list that was soon generated today provides new subjects for introduction into the study whenever vacancies occur.

Although Dr. Peter died suddenly on March 31, 1959, the recruitment system and the *esprit de corps* he helped to establish among the early participants have remained over the years. The study owes much to Dr. Peter and stands as a memorial to his vision. Many people—physicians, scientists, technicians, programmers, and secretaries—have contributed to the successful continuation of the study over the past 23 years. To mention all of them by name would require an additional volume of this book. To all of them, our sincere thanks.

Finally, acknowledgment must be made to the subjects themselves for their willingness to participate in procedures that were often demanding and tedious, and sometimes uncomfortable, in order to improve and extend knowledge about aging. Without their loyalty and dedication no study would have been possible.

Acknowledgments

The success and continuity of the Baltimore Longitudinal Study of Aging (BLSA) are due to the dedicated efforts of many people. First of all we must acknowledge the debt owed to the loyal subjects of the study, who faithfully met appointments and gave their maximum effort in performing tests of physiological and psychological functions. Our subjects accepted all kinds of tests as a challenge, even when some were boring and others uncomfortable. Without their loyalty and continued participation the study could not have endured for 23 years, and would not now look forward to a productive future.

We gratefully acknowledge the role of Mrs. Elizabeth Strawn, Nursing Supervisor, Baltimore City Hospitals' (BCH) Chronic Disease Hospital, who supported the study by finding and making available resources that were essential to the early stages of the study.

We are also indebted to the staff of the Gerontology Research Center (GRC), to the many physicians who conducted the detailed histories and physical examinations, to the scientists who designed the study, and to the technical staff who conducted tests and measurements.

We are especially indebted to the late Mr. Arthur H. Norris, who, from the inception of the study in 1958 until his untimely death in December 1980, provided major scientific input, consummate administrative acumen, and above all genuine personal support to the BLSA participants both during and between their visits.

Many people took part in the development of the collaboration between the National Institutes of Health (NIH) and the BCH, which resulted in the formation of the GRC on the grounds of the BCH. Dr. Henry Sebrell, Director of NIH, Dr. James Watt, Director of the National Heart Institute (NHI), and Dr. Robert Berliner, Scientific Director of NHI, played key roles over the years. Dr. T.J.S. Waxter, Director of Public Welfare, Baltimore City, Mr. Francis Davis, Chairman of the Welfare Board of Baltimore, Mr. P.J. McMillin, Superintendent of BCH, and Dr. John T. King, Chief of Medicine, BCH, were instrumental in the development of the GRC, which is essential to the BLSA.

Art-work for the illustrations in this volume was prepared for publication by the Photography and Arts Section of the GRC, under the supervision of Mr. Paul R. Ciesla.

Especial thanks are due to Ms. Paula Wernick and Ms. Eloise Acord for editorial, bibliographic, and manuscript-processing support without which the work on the book could not have proceeded as smoothly as it did. Edwin Watkins shepherded the book through writing, editing, design, and production.

It is also important to acknowledge the role of the GRC itself in providing a stable and visible "home" for the BLSA, as well as resources that include housing facilities for subjects, laboratories, and a staff committed to the study of aging.

Contents

page

INTRODUCTION 1

CHAPTER I. Methods for the Study of Aging 5

The Cross-Sectional Method and Its Limitations 5
Advantages and Limitations 5
Effects of Differential Mortality 6
Birth-Cohort Effects 8
Disease Effects 9
The Longitudinal Method 9
Advantages of the Longitudinal Method 10
Age Regressions for Individual Subjects 10
Predicting Outcomes 10
Continuity of Study Population 13
Operational Challenges in Longitudinal Studies 14
Recruitment and Screening of Subjects 14
Attrition 15
Expansion of Subject Panel 15
Strategies of Analytical Design 15
Maintaining Uniformity of Methods and Quality Control 16
Data Storage and Retrieval 17
Staffing 17
Financing 17

CHAPTER II. Longitudinal Studies: Past and Present 19

Introduction 19
Studies of Growth and Development 20
Studies of Selected Populations of Adults 20
The Cardiovascular Disease Project at the University of Minnesota 20
The Duke Studies 22
Normative Aging Study 24
The 1000-Aviator Study, Pensacola 26
The National Institute of Mental Health Study 28
The Basel, Switzerland, Study 30
The Bonn, West Germany, Study 33
Studies of Community Populations 34
The Framingham Study 34
The Tecumseh, Michigan, Community Health Study 35
The Atomic Bomb Casualty Commission Study 36
Studies of Specific Variables 39
Intellectual Functions 39
Physiological Variables 40
Summary 42

page

CHAPTER III. Design and Operation of the Baltimore Longitudinal Study of Aging 45

- History 45
- Objectives 46
- Operation of the BLSA 47
  - Administration 47
  - Participants 48
  - The Tests 55
- Approaches to Interpretation of Data 57
  - The Aging-Disease Dilemma 57
  - The Concept of Clinical "Clean-Up" 59
- The Data-Management System 60

CHAPTER IV. Tests Administered 63

- Introduction 63
- Clinical Evaluation 63
  - History 63
  - Physical Examination 63
  - Laboratory Tests 69
  - The Cornell Medical Index Health Questionnaire 69
- Genetic Characteristics 70
  - Family History 70
  - Genetic Markers 70
- Body Structure and Composition 70
  - Anthropometry 70
  - Bone-Mineral Measurements 70
  - Subcutaneous Fat 71
  - Estimates of Components of Body Mass 71
  - Strength Tests 72
  - Basal Metabolism 72
- Nutrition 72
  - Dietary Habits 72
  - Nutrition History 73
  - Pyridoxine (Vitamin-B_6) Status 73
- Neuromotor Function and Exercise 73
  - Tapping Test 73
  - Oral Motor Evaluation 73
  - Reaction Time 73
  - Nerve-Conduction Velocity 74
  - Physical Activity: Muscular Strength and Work 75
  - Physical Activity: Oxygen Uptake 75
- Renal Function 78
  - Creatinine Clearance (24-Hour) 78
  - Urinalysis 78
  - Concentrating Ability of the Kidney 78

page

Pulmonary Function 78
  Spirometry 78
  Maximum Breathing Capacity 79
  Pulmonary Gas Distribution 79
  Lung and Chest-Wall Compliance 79
  Smoking History 80
  Chest X-Ray 80
Cardiovascular Function 80
  Screening for Cardiovascular Disease 80
  Evaluation of the Effect of Age on Cardiovascular Function 82
Carbohydrate Metabolism: The Glucose-Insulin Homeostatic System 84
Metabolism of Drugs 86
  Antipyrine Metabolism 86
  Ethanol Metabolism 86
Hypothalamic-Pituitary Function 87
  Reproductive Hormone System 87
  Hypothalamic-Posterior Pituitary Function 87
The Immune System 88
Cell Replication 88
Special Senses 89
  Eye Tonography 89
  Visual Screening Tests 89
  Fundus Photography 89
  Audiometry 89
Cognitive Performance 89
  Intelligence Tests 89
  Learning, Memory, and Decision Tasks 90
  Problem Solving 91
Personality and Developmental Characteristics 91
  Personality 91
  Developmental Antecedents 92
Stress and Coping Processes 93
  Stress 93
  Coping and Defense Mechanisms 93
  Adaptational Outcomes 93
Marital and Sexual Experience 94

CHAPTER V. Cross-Sectional Studies of Aging in Men 95

Physiology 95
  Cell-Culture Senescence and *In-Vitro* Life Span 95
  Body Composition 97
  Nutrition 105
  Drug Metabolism 112
  Pulmonary System 114
  Cardiovascular System 117

page
Cardiac Performance and Aging: Occult Disease 117
Cardiac Function at Rest 118
Cardiovascular Response to Stress 121
Summary: Hemodynamic Function at Rest and in Response to Stress 125
Electrophysiology 126
Renal System 127
Endocrine System 128
Immune System 137
Oral/Dental 148
Behavior 148
Psychophysiology 148
Effect of Age on Reaction Time and Vigilance 148
Electroencephalographic (EEG) Studies 151
Electrodermal Activity 154
Time Perception and Psychomotor Function 156
Cognitive Functions 158
Learning and Memory 158
Equivalence of Information in Concept Identification 161
Daydreaming 162
The Marital and Sexual History 164
Activities and Attitudes 166
Coping with Stressful Events 166

CHAPTER VI. Longitudinal Studies of Aging 169

Introduction 169
Conceptual Considerations in the Identification of Age Changes 169
Body Size and Composition 171
Fat Redistribution and Changing Body Dimensions 171
Height and Weight 173
Joint Degeneration in Osteoarthritis of the Hand 173
Physiological Performance 174
Basal Metabolism 174
Renal Function 175
Serum Cholesterol 176
Cardiovascular Function 177
Cognitive Performance 179
Vigilance 179
Problem Solving 180
Learning 182
Visual Memory 183
Temperament as a Predictor of Change in Visual Retention 184
Blood Pressure and Intelligence 185
Personality Characteristics 187
Age and Stability of Personality 187
Historical Effects on Test Data 192

page
Eliminating Response Bias as an Explanation for Stability 194
The Stability Model: Implications for Gerontology and Geriatrics 196
Longitudinal Changes in Adjustment to Aging 197
Personality and Adjustment to Aging 197
Somatic Complaints as a Function of Age and Neuroticism 198
Hypertension, Somatic Complaints, CAD, and Personality 199
Research on Personality and the Life Course 202
Outcome Studies 202
Obesity and Longevity 202
Age and Some Physiologic Variables 203
Predicting Coronary Events in Asymptomatic Subjects 204

CHAPTER VII. BLSA: Acta and Agenda 207

Introduction 207
Individuality and Specificity of Aging 207
Patterns of Change with Age 208
Future Directions 209

APPENDIX. Longitudinal Studies in the BLSA Population App.1

Arenberg, D., A Longitudinal Study of Problem Solving in Adults 5

Arenberg, D., Differences and Changes with Age in the Benton Visual Retention Test 14

Arenberg, D., Estimates of Age Changes on the Benton Visual Retention Test .. 21

Arenberg, D., Changes with Age in Problem Solving 25

Arenberg, D., Memory and Learning Do Decline Late in Life 40

Borkan, G.A., and Norris, A.H., Fat Redistribution and the Changing Body Dimensions of the Adult Male 51

Costa, P.T., Jr., Fleg, J.L., McCrae, R.R., and Lakatta, E.G., Neuroticism, Coronary Artery Disease, and Chest Pain Complaints: Cross-Sectional and Longitudinal Studies 71

Costa, P.T., Jr., and McCrae, R.R., Somatic Complaints in Males as a Function of Age and Neuroticism: A Longitudinal Analysis 79

Costa, P.T., Jr., and McCrae, R.R., An Approach to the Attribution of Aging, Period, and Cohort Effects 92

Costa, P.T., Jr., and McCrae, R.R., Concurrent Validation after 20 Years: The Implications of Personality Stability for Its Assessment 105

Costa, P.T., Jr., and McCrae, R.R., Personality as a Lifelong Determinant of Well-Being 129

Costa, P.T., Jr., McCrae, R.R., Andres, R., and Tobin, J.D., Hypertension, Somatic Complaints, and Personality 147

Costa, P.T., Jr., McCrae, R.R., and Arenberg, D., Enduring Dispositions in Adult Males 163

page

Costa, P.T., Jr., McCrae, R.R., and Arenberg, D., Recent Longitudinal Research on Personality and Aging App.171

Costa, P.T., Jr., McCrae, R.R., and Norris, A.H., Personal Adjustment to Aging: Longitudinal Prediction from Neuroticism and Extraversion 215

Costa, P.T., Jr., and Shock, N.W., New Longitudinal Data on the Question Whether Hypertension Influences Intellectual Performance 223

Douglas, K., and Arenberg, D., Age Changes, Cohort Differences, and Cultural Change on the Guilford-Zimmerman Temperament Survey 234

Elahi, V.K., Elahi, D., Andres, R., Tobin, J.D., Butler, M.G., and Norris, A.H., A Longitudinal Study of Nutritional Intake in Men 245

Ensor, R.E., Fleg, J.L., Kim, Y.C., deLeon, E.F., and Goldman, S.M., Longitudinal Chest X-Ray Changes in Normal Men 264

Fleg, J.L., Das, D.N., and Lakatta, E.G., Right Bundle Branch Block: Long-Term Prognosis in Apparently Healthy Men 272

Hershcopf, R.J., Elahi, D., Andres, R., Baldwin, H.L., Raizes, G.S., Schocken, D.D., and Tobin, J.D., Longitudinal Changes in Serum Cholesterol in Man: An Epidemiologic Search for an Etiology 278

McCrae, R.R., and Costa, P.T., Jr., Aging, the Life Course, and Models of Personality 292

McCrae, R.R., Costa, P.T., Jr., and Arenberg, D., Constancy of Adult Personality Structure in Males: Longitudinal, Cross-Sectional, and Times-of-Measurement Analyses 304

Norris, A.H., Mittman, C., and Shock, N.W., Changes in Ventilation with Age . 311

Plato, C.C., and Norris, A.H., Osteoarthritis of the Hand: Longitudinal Studies . 318

Potter, J.F., Elahi, D., Tobin, J.D., and Andres, R., Effect of Aging on the Cardiothoracic Ratio of Men 325

Quilter, R.E., Giambra, L.M., and Benson, P.E., Longitudinal Age Changes in Vigilance over an Eighteen-Year Interval 331

Robertson-Tchabo, E.A., Arenberg, D., and Costa, P.T., Jr., Temperamental Predictors of Longitudinal Change in Performance on the Benton Revised Visual Retention Test among Seventy-Year-Old Men: An Exploratory Study 335

Rowe, J.W., Andres, R., Tobin, J.D., Norris, A.H., and Shock, N.W., The Effect of Age on Creatinine Clearance in Men: A Cross-Sectional and Longitudinal Study 345

Schlesselman, J.J., Planning a Longitudinal Study. I. Sample Size Determination 354

Schlesselman, J.J., Planning a Longitudinal Study. II. Frequency of Measurement and Study Duration 362

Shock, N.W., Energy Metabolism, Caloric Intake, and Physical Activity of the Aging 372

page
Shock, N.W., Andres, R., Norris, A.H., and Tobin, J.D., Patterns of Longitudinal Changes in Renal Function App.384
Tobin, J.D., Physiological Indices of Aging 387
Tzankoff, S.P., and Norris, A.H., Longitudinal Changes in Basal Metabolism in Man 396
References 401
Index of Authors Cited 419
Index 425

Tables

page

I.1 A Simple Cross-Sequential Design 16
I.2 A Simple Time-Sequential Design 16
II.1 Duke Study. I. Longitudinal Observations: Summary of Variables 22
II.2A Duke Study. II. Adaptation Study Design 23
II.2B Duke Study. II. Summary of Variables 24
II.3 Normative Aging Study (Boston VA Outpatient Clinic) 25
II.4 The 1000-Aviator Study 27
II.5 The 1000-Aviator Study: Summary of Tests Administered 28
II.6 The National Institute of Mental Health Study: Interviews and Tests Administered 29
II.7 The National Institute of Mental Health Study: Changes in Survivors after 11 Years 31
II.8 The National Institute of Mental Health Study: Significant Differences between Survivors and Nonsurvivors 32
II.9 The Bonn Longitudinal Study of Aging: Number of Participants and Dropouts and Number of Control-Group Participants 33
II.10 The Framingham Study 35
II.11 The Tecumseh Study: Tests Administered 36
II.12 Atomic Bomb Casualty Commission—Hiroshima Study: Tests Administered to Clinical Subsample 37
II.13 Selected Longitudinal Studies of Adults 43
III.1 Recruitment Pattern by Age, Cycles A-K 49
III.2 Panel Age Composition, Cycles A-K 50
III.3 Cumulative Number of Male Subject Visits by Age at First Visit 50
III.4 Percentage Distribution of Male Subjects by Age and Socioeconomic Characteristics at First Visit 52
III.5 Percentage Distribution of Male Subjects by Age and Personal Attributes at First Visit 53
III.6 Percentage Distribution of Male Subjects by Mean Age and Selected Attributes at Each Cycle 54
III.7 Follow-up Study of Male Subjects Seen at Least Once from February 1958 to June 1977 55
III.8 Number of Subjects Excluded from Normal Group by Disease Category and Age Group 58
IV.1 Tests Administered 64

page

V.1 Characteristics of Skin Explants and Cell Cultures Derived from Young and Old Human Donors 96

V.2 Calories Expended during Various Physical Activities 105

V.3 Total Daily Intakes of Various Nutrients in Men of Different Ages 107

V.4 Mean Values, Range of Values, and Standard Deviations of the Distributions of Lung-Clearance Index by Age Groups 116

V.5 The Prevalence of Coronary Artery Disease in the Elderly as Assessed by Resting and Stress Criteria 117

V.6 Effect of Adult Age on Resting Cardiac Function 120

V.7 The Effect of Age on Hemodynamic Performance at Maximum Exercise in BLSA Subjects 123

VI.1 Six- and 12-Year Retest Coefficients for GZTS Scales for Total Samples and 12-Year Retest Coefficients for Three Age Groups... 190

VI.2 Correlations between GZTS Scales Administered 1959-1966 and NEO Facets Administered in 1980 193

VI.3 T-Scores for Live and Dead Groups 204

Figures

Dr. W.W. Peter v

I.1 Confounding Effect of Selective Mortality on Inferences About Age Changes 6

I.2 Confounding Effect of Selective Mortality on the Magnitude of Age Changes 7

I.3 Confounding Effect of Selective Mortality Masking Age Changes 8

I.4 Height Curves during Adolescent Growth 12

I.5 Basal Heat Production 12

I.6 Menarche and Basal Heat Production 13

III.1 BLSA Volunteers 48

IV.1 Hand-Grip Strength: Smedley Dynamometry 76

IV.2 Maximum Work by Bicycle Ergometry 77

IV.3 Treadmill Stress ECG 81

V.1 Basal Oxygen Consumption 98

V.2 24-Hour Creatinine Excretion 99

V.3 Nonmuscle Oxygen Consumption 100

V.4 Lean Body Mass and Creatinine Excretion 101

V.5 Forearm Fat Thickness and Limb Diameter 102

V.6 Calf Fat Thickness and Limb Diameter 103

V.7 Daily Intake of Calories, Calcium, Iron, Vitamin A, Ascorbic Acid, and Thiamine 108

V.8 Percentage of Caloric Intake Derived from Fat, Carbohydrate, and Protein 109

V.9 Daily Caloric Intake, Basal Metabolic Rates, Energy Expenditures, and Energy Expenditures per Unit of Body Weight 110

V.10 Plasma Pyridoxal Phosphate Levels 111

V.11 Ethanol Concentration in Blood Water 113

V.12 Peak Blood-Ethanol Concentration at End of Infusion Period 113

V.13 Lung Volumes Divided by Body-Surface Area Plotted Against Age 114

V.14 Maximal Breathing Capacity Divided by Body-Surface Area Plotted Against Age 115

V.15 Relationship between Age and Diastolic Cavity Dimension 118

V.16 Relationship between Age and Ejection-Fraction Index 119

V.17 Age-Related Increase in Left-Ventricular Wall Thickness 121

V.18 Relationship between Age and Rate of Left-Ventricular Early Diastolic Filling 122

page
V.19 Plasma-Norepinephrine Concentration during Treadmill Exercise 124
V.20 Increase in Heart Rate Induced by Isoproterenol 125
V.21 Age-Adjusted Percentile Rank in Creatinine Clearance 127
V.22 Cortisone-Glucose Tolerance Test 129
V.23 Age-Adjusted Percentile Ranking in Cortisone-Glucose Tolerance Test 130
V.24 Age Effect on Response to Tolbutamide 131
V.25 Age-Adjusted Percentile Rank on Tolbutamide Response Test 133
V.26 Serum Osmolality and AVP Concentration 136
V.27 Binding of (——) ^{3}H-dihydroalprenolol to Mononuclear Cell Membranes 136
V.28 Counts of White Blood Cells, Segmented Neutrophiles, Lymphocytes, and Monocytes 138
V.29 Functional Characteristics of T Lymphocyte and B Lymphocyte 138
V.30 Percentage of Blood Lymphocytes with Membrane Immunoglobulin .. 139
V.31 Percentage of Blood Lymphocytes That React with Monoclonal Antibodies 140
V.32 Bactericidal Capacity of Granulocytes 141
V.33 Superoxide Generation by Granulocytes 142
V.34 Numbers of Immunoglobulin-Producing Cells Induced by Pokeweed Mitogen Stimulation 143
V.35 Numbers of Immunoglobulin-Producing Cells in Replicate Cultures of B Cells from Responders and Nonresponders 144
V.36 Geometric Mean Percentage of ^{51}Cr Release According to Age and Sex 144
V.37 Ability of T Lymphocyte to Divide 145
V.38 Phytohemagglutinin Stimulus Responses 145
V.39 Changes in Number of High-Space and Low-Responder Cultures 146
V.40 Serum-Immunoglobulin Determinations 147
V.41 Reaction Time as a Function of Stimulus Information 149
V.42 Decline in Vigilance 150
V.43 Reaction Time and Period of EEG 151
V.44 Age and Period of EEG 152
V.45 Age and Strength of Arm and Shoulder Muscles 158
V.46 Age and Maximum Work Rate for Four Cranking Rates 158
V.47 Verbal Learning by Age-Pace Groups 160
VI.1 18-Facet NEO Model 188

Normal Human Aging:

The Baltimore Longitudinal Study of Aging

Note: References that appear in italics in the text identify longitudinal studies on the BLSA population. These studies are summarized in Chapter VI and, unless they are still in press, are reprinted in the Appendix.

Introduction

Since the inception of the Baltimore Longitudinal Study of Aging (BLSA) in 1958, its intent has been to trace the effects of aging in humans. It attempts to distinguish between the true effects of aging and those processes, including disease, socioeconomic disadvantage, and lack of educational opportunity, that may also appear or become more pronounced with time but are biologically irrelevant to the underlying mechanisms of human aging. It has sought to achieve this goal by making measurements on more than 1000 male volunteers[1] who at their entry into the study were aged 17 to 96 years and were living independently in the community. This report of its first 23 years of activity is an initial compilation of methods and findings pointing the way to further research and analysis that will enhance our understanding of the complex and highly individual processes of human aging.

Changes that take place during the early stages of life, which are almost entirely reflections of growth and development, are usually characterized by increases in size and in complexity of structure and function. By contrast, aging is manifest in a series of physiological and behavioral changes that occur after the attainment of maturity in all members of a species. Although aging changes proceed at a much slower rate than changes that occur with growth and development, and are often associated with decrements in performance, it cannot be assumed that aging is always characterized by impairment of function. In many people, for example, vocabulary continues to expand throughout life. Aging in adults is also characterized by great diversity; old organisms show a greater range of variation in many physiological and psychological indicators than any other adult age group. There are extraordinarily "young" 80-year-olds, along with extraordinarily "old" 40-year-olds.

Although the incidence of disease increases with age, aging and disease are not synonymous. Aging is a normal concomitant of the passage of time that takes place in everyone; disease occurs in only a part of the population. The changes by which aging is manifest, at all levels including the cellular, reduce an individual's adaptive capacities and the speed and excellence of his performance, and increase his susceptibility to disease and pathological processes. As a result, a traumatic event such as an accident or exposure to a disease, which may be of minor consequence to a young individual, may be of much greater consequence, even fatal, in the elderly. As adaptability and reserve capacities diminish, mortality rates increase. Although there are wide individual differences in the rate at which age changes take place, aging affects all members of a population, while specific diseases and accidents are selective.

Since many of the changes associated with aging reflect a reduction in adaptability and performance that may also characterize specific diseases, the effects of aging and disease are difficult to separate. One approach is to examine all subjects in great detail and to exclude from the analysis of age trends data from any subjects who show evidence of pathology or disease. Although this procedure may eliminate subjects with gross pathology, there is no assurance that subjects with the early stages of a disease have been identified. The development of new diagnostic techniques more precise and sensitive than those previously available may reveal that some subjects, formerly judged normal, were in fact already in the early stages of a disease. It is thus impossible to be sure that only "healthy" subjects have been included; it can only be stated that individuals with pathology detectable by the best methods then available had been eliminated at that stage of data analysis.

---

[1] Since January 1978, women have been systematically added to the BLSA. As of June 30, 1981, more than 300 had been examined and tested at least once, 150 two or more times. It is anticipated that the ultimate number of female BLSA participants will approach 700. Because of the short time over which observations have been made, no longitudinal data on aging in women will be reported in this volume. However, certain cross-sectionally derived data are reported in Chapter V.

The BLSA has attempted to measure age changes by making serial measurements on individual members of a group of community-residing adults. In order to minimize the effects of socioeconomic and educational factors on test results, subjects were recruited from a well-educated population with above-average income, which also had access to good medical care. Careful medical examination for specific conditions or diseases that might influence the "normality" of a particular performance has been employed to exclude "abnormal" data. As a result, age differences and age changes observed in this highly selected group of participants may be the best available index of "pure," i.e. optimal, aging.

Studies of growth and development, or of aging during the early part of life, have traditionally been conducted by the cross-sectional method, in which subjects of various ages are measured simultaneously. The effects of aging or growth were thus inferred from *differences* between average values found in groups of subjects of various ages. Only the longitudinal method, in which serial measurements are made on the same subjects as they age, can identify age *changes.*

This book is addressed to those who are interested in the objectively measurable changes that take place in adults as they grow older. The tests used to measure different aspects of aging, the functions they evaluate, and the methodology used are described. Neither the theory of longitudinal studies nor the mathematical and statistical issues inherent in longitudinal data analysis are discussed in detail, since other books and articles cover these issues (Jones, 1958; Birren and Renner, 1977; Schaie, 1977; Baltes, 1968; *Schlesselman, 1973a,b;*[2] Schulsinger et al., 1981). Although a description of past and existing longitudinal studies on adults is necessary to answer the primary question—"Why another longitudinal study?"—the book does not include an exhaustive review of all the longitudinal studies that have been conducted on children and adults. The literature review in Chapter II is limited to studies that illustrate the strengths and weaknesses of the longitudinal method and those that were conducted on normal community-residing adults—studies which may reasonably be compared with the BLSA.

The general design of the study, characteristics of the subjects, the tests used, data storage and retrieval, and methods of data analysis are described in Chapter III. Chapter IV describes in detail the testing procedures followed.

The results of published BLSA cross-sectional analyses are summarized in Chapter V. Those based on longitudinal examinations of selected variables (anthropometry, physiological performance, cognitive performance, and personality characteristics) are summarized in Chapter VI, and the articles themselves, except a few that are still in press, are reprinted in the Appendix, since they provide models that may be followed for the longitudinal analysis of other variables.

---

[2] References that appear in italics in the text indicate longitudinal studies on the BLSA population. These studies are summarized in Chapter VI and, unless they are still in press, are reprinted in the Appendix.

The primary focus of the BLSA is the scientific characterization of aging in individuals over the entire adult life span. Unlike many other studies whose chief purpose was to identify factors that increase the probability of the development of specific diseases, principally cardiovascular, the BLSA has from its inception emphasized healthy aging; it thus includes a much broader spectrum of both physiological and psychological tests than other studies. In studies of aging it is particularly important to rule out the possibility that occult disease is the source of the observed differences. At the same time, no subject in whom disease was discovered during the study was excluded or dropped, although observations made on such individuals were excluded from analyses for pure age effects. As a result, the progressive effects of such conditions as diabetes and cardiovascular disease on specific physiological functions can be described.

The BLSA was extraordinarily fortunate in its early recruitment and retention of a highly homogeneous group of well-educated and dedicated volunteer participants who were willing to commit themselves to visit the Gerontology Research Center (GRC) at regular intervals for an indefinite period, perhaps for the remainder of their lives. Without the initiative and assistance of Dr. W.W. Peter in their recruitment, and without the continued commitment of the participants (see "Dedication," "Acknowledgments," and Chapter III), the essential stability of the sample population could not have been maintained. Because of their dedication, the study was able to collect reliable historical and background data, to compare measurements of responses to standardized stimuli under basal conditions with previous measurements, and to include more tests across a broader spectrum of research areas, as well as to characterize its subjects in much greater detail, than most other studies. Their loyalty has also made it possible to introduce many tests that not only are time-consuming but often involve some degree of personal discomfort.

Few studies have been pursued over periods long enough, and with enough repeated measurements, to permit true longitudinal analysis of changes in a variable in a single individual as he ages. The BLSA is unique in providing as many as 21 sequential measurements on the same subject, from which linear regressions on age and standard errors of estimate have been calculated for some functions. As of June 30, 1981, five or more data points were available for some measurements on 667 subjects.

The BLSA is unique both in the frequency of visits and in the time devoted to each. A visit of 2½ days, which includes two nights at the GRC, makes possible measurement under basal conditions of many physiological functions, such as blood pressure, heart rate, and oxygen consumption, as well as such time-consuming tests as studies of glucose and insulin homeostasis, hypothalamic-pituitary-gonadal function, and 24-hour renal-clearance rates. Measurements made both under basal conditions and during stress are necessary to identify more subtle changes in function and to detect occult disease. This is particularly true of the cardiovascular system; clinically occult coronary artery disease, which is present in a substantial proportion of elderly subjects, can be detected only through the imposition of stress on the heart. The length and frequency of visits have made possible a large number of measurements for many participants, with a consequent increase in the statistical reliability of mean values and regression coefficients.

The BLSA is also unique in the potential it offers of future study in its present and projected population. Participants recruited in their early adult years are still actively committed to the study, while new subjects are introduced in order to maintain the population and to make possible analyses that will help identify "period"[3] and birth-cohort effects. Although it is clear from the BLSA experience to date that both cross-sectional and longitudinal analyses are required for studies of aging, neither method by itself identifies exclusively the effects of aging. In cross-sectional analysis, age differences between groups of subjects are confounded with birth-cohort differences, i.e., effects due to events that occur at different ages in the lives of persons born at different times. Some of these events, such as economic depressions, war, or shifts in social policy, may have far-reaching effects that vary with the subject's age and situation at the time the events occurred. Longitudinal analysis confounds changes due to aging with those that result from "period" or secular influences, which affect whole populations. The attempt to differentiate among the three effects is strengthened by the addition of "cross-sequential" and "time-sequential" analyses described in Chapter I. Although in many areas of investigation this sort of analysis remains to be done, the data bases necessary for its accomplishment are now becoming available.

The BLSA has already achieved some of the goals set for the study. It has contributed to knowledge about aging in normal persons and has shown that age changes can be estimated for some variables in individual subjects. It has highlighted, and suggested approaches to, some of the problems inherent in the analysis of both cross-sectional and longitudinal data for age changes.

A number of critical questions remain for the future, among them the interrelations among age changes in different organ systems, the time courses of different diseases, and the effects of critical life events. In the coming years the study on women will mature, and insight should be gained into the remarkable but still unexplained difference between the sexes in disease development and longevity.

---

[3] See Chapter I for definitions of terms.

CHAPTER I

Methods for the Study of Aging

The two principal methods by which the effect of aging on a variable can be measured are the cross-sectional and longitudinal designs. The cross-sectional method is characterized by measurements made at approximately the same time on a large number of subjects covering the entire adult age span. Age changes are not measured directly but are inferred from the differences in mean values observed in different age groups or from the overall regression of the measurement on age. Only average differences between age groups are identified.

The longitudinal method is characterized by serial measurements of a specific variable on the same subject as aging occurs; it thus identifies age changes in individuals in addition to average differences between groups of subjects of different ages. Since each method has its strengths and weaknesses, the quantitative measurement of aging requires the application of both. Other variations of the two ("cross-sequential" and "time-sequential" approaches) are needed to control factors that cannot be isolated by either method alone.

THE CROSS-SECTIONAL METHOD AND ITS LIMITATIONS

1. Advantages and Limitations

Growth and development have traditionally been studied by the cross-sectional method, by which the average values of a variable are calculated for groups of subjects distributed according to age. Growth is inferred from the progressive increase of the average values for height or weight in groups of growing children; that is, the regression of measurements on age is viewed as an index of growth rate. Such analysis neither provides a direct measurement of age changes nor specifies the magnitude or rates of change in individual subjects. Its primary advantage is that the presence of age trends in a group of subjects can be detected fairly quickly. Caution is necessary in its interpretation, however, since differences between age groups include birth-cohort[1] as well as age effects (see below, "Strategies of Analytical Design").

Students of child development recognized this limitation of cross-sectional analyses in the early part of this century and initiated longitudinal studies, in which measurements of height and weight were repeated at short intervals to generate growth curves for individual subjects (Dearborn et al., 1938). Although important findings have emerged, the longitudinal method has been used in only a few studies of adults (see Chapter II), because of the difficulties of recruiting and retaining subjects for repeated measurements over long periods of time, as well as of finding the necessary long-term financial support.

---

[1] A birth cohort consists of individuals born in the same arbitrarily chosen interval of time. Since specific environmental conditions occur at different ages for subjects from different birth cohorts, the effects of such events may be confounded with the effects of aging in their influence on cross-sectional measurements.

2. Effects of Differential Mortality

As a study population ages it becomes more and more selected, since death occurs more frequently among old than among young subjects. By age 70, the population available for study represents only about 50% of the original birth cohort. Averages derived from measurements in young adults are thus based on observations of subjects some of whom will not live to age 70, while data from older subjects obviously represent an "élite" population that has survived.

Age differences in a measured variable do not necessarily reflect changes in individuals or even the average changes in specific age groups, since deaths do not occur randomly throughout the population but are more likely among individuals whose characteristics increase their susceptibility. The selective nature of the process is implied by the term "differential mortality." This effect increases with age.

The influence of differential mortality on inferences about age changes made on the basis of cross-sectional measurements of age differences can be visualized from the theoretical curve in Figure 1. Consider the hypothetical variable X whose level varies among individuals but does not change with aging in any individual. Each of the horizontal lines in the figure represents a single individual as he ages. Suppose further,

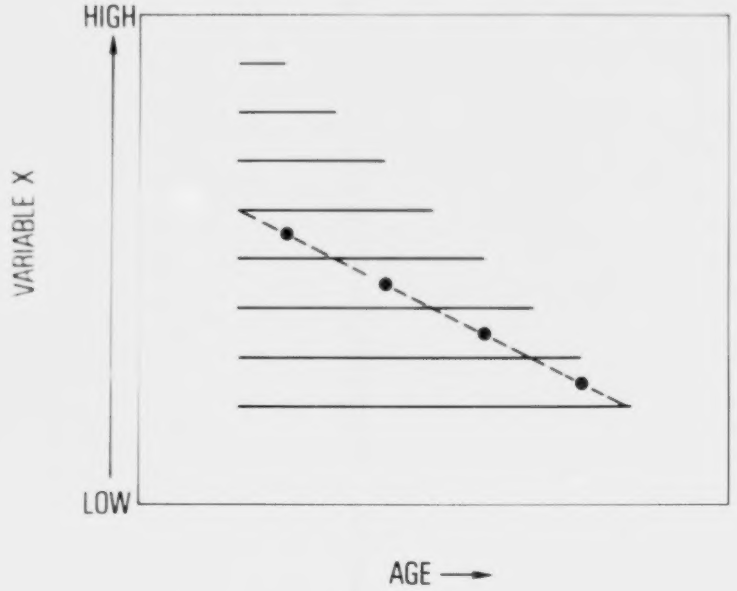


Figure I.1. Confounding effect of selective mortality on inferences about age changes. Each solid line segment represents the pattern of change in an individual (in this case there is no change with aging, and death occurs at the end of the line). High values for the variable X are assumed to be deleterious. The closed circles represent mean values that would be obtained in a cross-sectional study; the dashed line connecting these dots would then correctly represent age *differences* among age groups, but the inference that age *changes* were occurring in individuals would be erroneous.

as the figure shows, that a high value of X is deleterious, so that deaths occur first in high-X and last in low-X subjects. As the figure indicates, the mean value for X at the earliest ages will be high, since all subjects are alive; at each succeeding decade the mean will fall, since the subjects with the highest values will have been eliminated by death. As a result, the circles connected by the dashed line represent the average values that would be obtained from a single cross-sectional study. The average values of X clearly fall with advancing age, although within individual subjects X does not change with age.

Figure 2 illustrates another way in which cross-sectional analysis may lead to erroneous conclusions. One can imagine a function (variable Y) that declines linearly across age groups but in which a floor effect, or "lethal limit," appears. In this case a cross-sectional study might show an average decline that would not accurately represent the magnitude of age changes in individuals. It is even possible (Fig. 3) to picture a variable that declines with age but in which younger subjects can tolerate lower levels than older subjects before dying. In this case the cross-sectional analysis might show no age differences at all despite a decline in function in individual subjects across the age span.

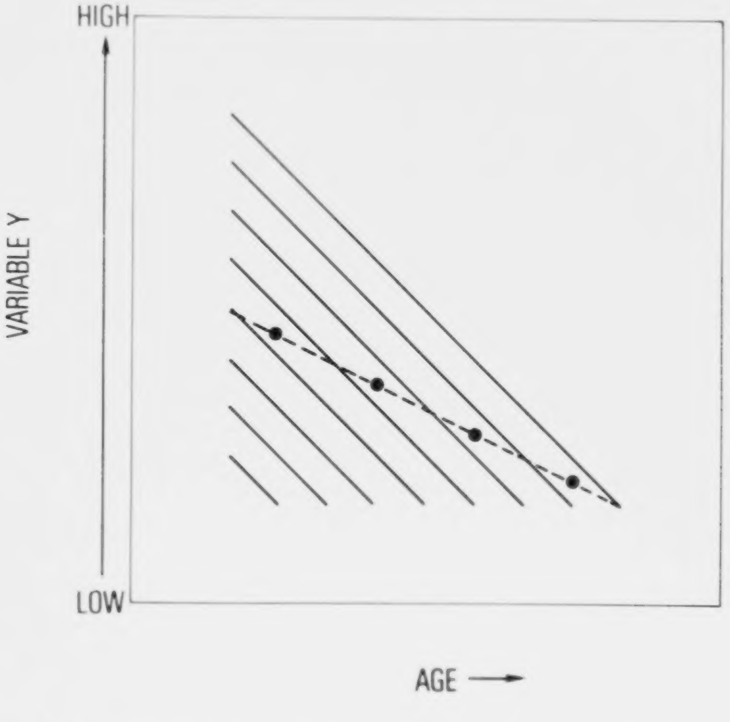


Figure 1.2 Confounding effect of selective mortality on the magnitude of age changes. See Fig. 1 for explanation of line segments. A floor effect or lethal limit is assumed for the variable Y. In this case the dashed line representing age *differences* would underestimate true age *changes*.

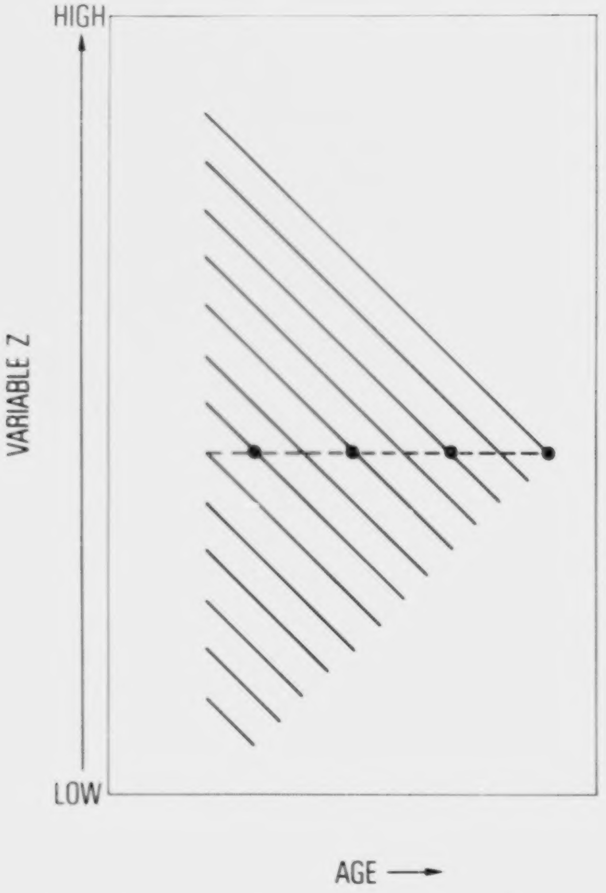


Figure I.3. Confounding effect of selective mortality as a result of which age changes are not revealed in cross-sectional studies. See Fig. 1 for explanation of line segments. A lethal limit that varies with age is assumed. In this case the dashed line would show no age *differences*, although large age *changes* in individuals had occurred.

3. Birth-Cohort Effects

Another limitation of cross-sectional studies is that young and old groups of subjects may differ in characteristics other than age that may also affect the measurements. Subjects born in a specified span of calendar years represent a birth cohort. An example of birth-cohort effects may be found in tests that are influenced by the level of education of the subjects. Most of today's young adults have completed high school, while a much smaller proportion of adults educated in the early years of this century reached that level of education. In any test in which level of education influences performance, young subjects will thus out-perform the old—but the difference may be due to education rather than to age.

Socioeconomic conditions early in life may also have affected older subjects differently from younger ones. For example, subjects who were 70 years old in 1980 were exposed to the effects of the economic depression of the 1930s when they were in their 20s, whereas subjects born after 1940 have not been subjected to such an event. Similarly, epidemics, wars, and other disruptions that occur at different points of the life cycle of different birth cohorts may influence test results in ways that the cross-sectional method cannot differentiate from true age changes (Birren and Renner, 1977).

4. Disease Effects

Since the occurrence of many diseases increases with advancing age, one of the primary problems in attributing differences between groups of subjects to aging is the necessity of excluding subjects suffering from diseases that influence the variable under study. This is an extremely difficult problem for which there is no certain remedy.

Many investigators, especially those concerned with behavioral and social research, have simply ignored the problem. Others have set criteria of health status, ranging from superficial to comprehensive, for inclusion in the study. These are commonly limited to the identification of a few specific disease states such as coronary artery disease or diabetes, or to arbitrary standards of normality in physical findings or laboratory tests such as blood pressure, blood glucose, or hemoglobin concentration. In many studies health status was determined by self-evaluation of the subjects; if a subject said he was in good health he was regarded as healthy. In only a few instances was a detailed physical examination carried out by physicians to screen subjects for the presence of specific diseases.

THE LONGITUDINAL METHOD

Some of the limitations of cross-sectional studies of aging can be minimized or overcome by a longitudinal design, in which the same subjects are measured repeatedly. The ideal longitudinal study of aging would provide observations on individuals over their entire life spans. Since this design is impractical, most longitudinal studies have been limited to specific periods of the life cycle. The phases of growth and development, for example, have received much more attention than adult aging. Among the many critical questions about how adults age that can be answered only from serial observations are the following:

Does the average curve of age differences based on cross-sectional data represent the average progression of aging in individual subjects?

How rapidly does an individual change with respect to a specific variable or test? That is, what is the rate of change? What is the diversity among individuals?

Is there a general aging factor, or does each organ system show a different pattern of aging? How are age changes in different variables related in individual subjects?

Do critical events in the life cycle of an individual affect aging? An answer to this question requires serial measurements in the subject before and after the event, which may take such forms as a heart attack or other severe illness, exposure to toxic substances or radiation, loss of job, retirement, loss of spouse, loss of mobility, or loss of independent living.

Can patterns or levels of performance at a given age be used as predictors of performance at a later age, or of longevity?

Can aging be distinguished from disease?

Does age influence the progression of such disease states as diabetes, arteriosclerosis, and hypertension?

Can a causal ordering be determined from serial observations when two variables are known to co-vary?

ADVANTAGES OF THE LONGITUDINAL METHOD

1. Age Regressions for Individual Subjects

The primary advantage of the longitudinal method is that it makes it possible to estimate age changes in an individual over a specified period, so that a "rate of aging" may be determined for any specified variable. The study must be so designed that enough observations are made in an individual to permit calculation of the standard error of estimate of the calculated regression on age. It is usually assumed that the regression of the variable on age is linear. Although the assumption of linearity may or may not be true, it is seldom possible to collect enough observations to reject it. A statistical analysis *(Schlesselman, 1973a,b)*[2] has been made of the experimental strategies (duration of the study, frequency of observations, number of observations) essential to achieve a specified reliability of estimate of the individual regression slopes for a specific variable. For one variable, equally satisfactory slopes might be computed by carrying out three tests in 14 years, ten tests in ten years, or 30 tests in six years. For another it might not be possible to compute individual slopes with satisfactory accuracy unless monthly examinations were carried out for 30 years. Two characteristics of the variable that necessitate different planning strategies are the mean rate of change with time (the mean regression on age for the population) and the degree of variance in the individual slopes. A third factor is the investigator's determination of the degree of accuracy required in the estimate of the age regression for an individual subject. The investigator may thus choose among many strategies for the design of a longitudinal study.

2. Predicting Outcomes

Events or processes experienced at various times may affect a person's health or functioning in later life, as well as survival. The longitudinal study design is valuable, sometimes essential, to identify events of significance and to quantify their long-term effects. In one sense, a true prospective longitudinal study with repeated periodic evaluations is not required to answer such questions: Information concerning past events may be obtained by history, and outcomes may be sought at a chosen point. The value of such an analysis, however, may be limited by the inaccuracies inherent in historical recall of distant events. The longitudinal approach, with reasonably frequent evaluations, decreases memory error, may provide objective evidence for the presence of an event, and permits more accurate identification of the time when both the event and its effects occur.

---

[2] References that appear in italics in the text indicate longitudinal studies on the BLSA population. These studies are summarized in Chapter VI and, unless they are still in press, are reprinted in the Appendix.

A major application of longitudinal data is the search for precursors and risk factors related to disease and death. The study of risk factors has traditionally been the province of epidemiological research. The distinctive characteristic of epidemiological research such as the Framingham study (see Chapter II) is its focus on the factors that influence the incidence of disease in a specified human population. One of the strengths of the BLSA lies in the wealth of its measurements, which permits intensive analysis of the antecedents of disease to a degree few epidemiological studies can match.

As a study of aging, the BLSA is also concerned with age-related stability or fluctuation in functions; the multidisciplinary data collected over a period of years can be used to determine predictors of change. Questions that may be addressed include: whether personality differences retard or accelerate declines in cognitive performance; whether regular exercise leads to better pulmonary functioning in old age; and what activities or attitudes contribute to an older person's sense of having lived a satisfying life.

A corollary of the search for significant predictive variables is the identification of their critical levels, the cut-points of which may vary with age. Although 24-hour creatinine clearance, for example, may show a large decline with age, its clinical significance depends on whether the decline predicts an increased likelihood of death or disease. The identification of a critical level or pattern of change may thus be a unique contribution of the longitudinal method.

In addition to assessing the impact of single variables on single outcomes, it is important to test combined effects of variables and complex outcomes. Thus, while myocardial infarction or death must be considered an outcome, more complex outcomes such as the ability to live independently of institutions, the ability to continue the activities of daily living, or the achievement of overall "successful aging" should be assessed. Even the rate of aging of a particular organ system may be analyzed as an outcome (the dependent variable) of other characteristics or risk factors earlier in life.

Studies on children illustrate the power of the longitudinal method in identifying the effects of events that occur at different chronological ages in different subjects. Figure 4A shows a series of curves of individual growth rates derived from serial measurements of height made on children as they aged from five to 18 years (Tanner, 1955). Each individual depicted achieved a maximum growth spurt at a different age, ranging from nine to 14 years. The mean of these curves obtained from cross-sectional measurements (shown by the dotted curve) grossly underestimates the magnitude of the individual adolescent growth spurts and fails to indicate the diversity of their timing. In Figure 4B the same curves have been so arranged that their points of maximum velocity coincide, and other points are plotted as deviations in time from that event. This method can be used to identify the effects of any event on other measurements, provided the time of occurrence of the event can be identified and a series of measurements taken before and after the event is available.

Longitudinal observations also provide an opportunity to identify the effects of physiological events on other variables—effects that cannot be identified by cross-sectional analysis (Shock, 1943). For example, cross-sectional observations of basal heat production in girls and boys aged 11.5 to 17.5 years lead to the conclusion that basal metabolic rate (BMR) falls gradually from age 12 to age 17.5 (Fig. 5) (Shock, 1942). However, when the data for girls are replotted with age of menarche as their zero point (Fig. 6), and serial observations taken at six-month intervals are plotted as

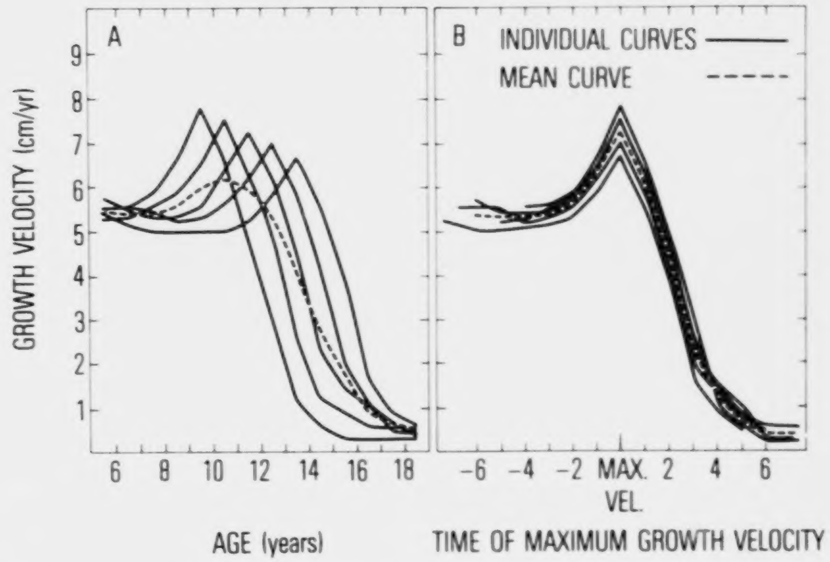


Figure I.4 Relation between individual and mean velocities during the adolescent growth spurt. A: The height curves are plotted against chronological age. B: The height curves are plotted as deviations from time of maximum growth velocity.
From Tanner (1955), after Shuttleworth (1939).

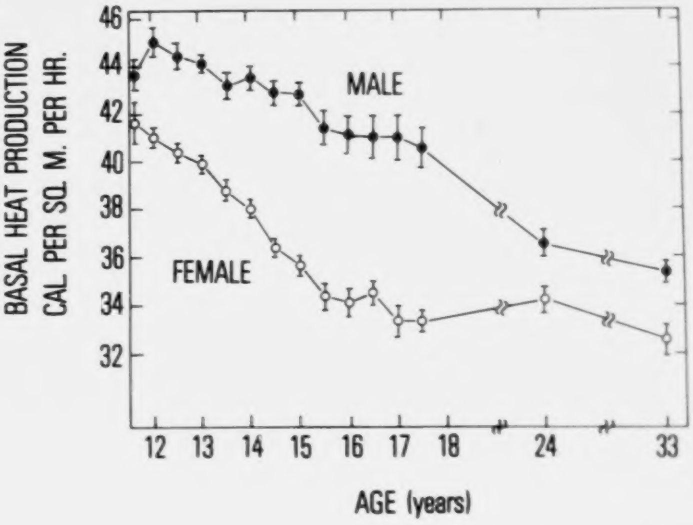


Figure I.5 Average basal heat production in males and females. Measurements made every 6 months from age 11.5 to 17.5 yr in same individuals. Points at ages 24 and 33 obtained from other subjects.
From Shock (1942, adapted).

deviations from age of menarche in each subject, it becomes apparent that menarche—a physiological event—is more important than chronological age in determining the adolescent fall in BMR.

The availability of serial observations also makes it possible to search for the effects of critical events in the life history of individuals—cessation of smoking, death of a spouse, retirement, loss of mobility, need for institutionalization—by comparing measurements made before the event with those made afterward.

3. Continuity of Study Population

The continued availability of a longitudinal population such as that of the BLSA offers unique opportunities for multidisciplinary investigations of the relations between aging and other variables, without the recurring need to recruit new subjects. A major advantage of the BLSA is that it has provided a study population in which thorough and repeated clinical evaluations have been carried out on all subjects. The resultant clinical records add efficiency to the overall operation of the study and provide essential background information on health status to other investigators whose primary interests may lie in domains other than clinical medicine.

Furthermore, when multiple variables are assessed a multiplicative effect occurs, in that each scientist is able to take advantage of the information generated by his colleagues to improve his own subject characterization and research interpretation. As data accumulate in the longitudinal study, not only the quantity but also the quality of the information available on each subject is thus increased.

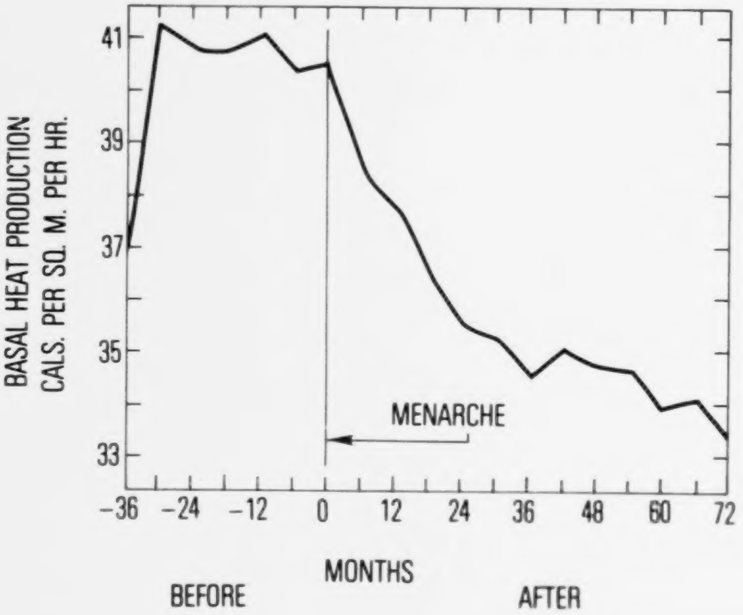


Figure I.6. Effect of menarche on basal heat production. Average values calculated for 6-mo intervals before and after menarche. Zero time for each girl is the age at which she first menstruated.
From Shock (1943).

An inevitable consequence of the existence of a longitudinal panel of subjects in an academic or research environment is that other scientists are attracted by the obvious advantages. This leads not only to a number of spin-off cross-sectional studies but also to the addition of new variables to be studied longitudinally. Any longitudinal study must first go through a phase of cross-sectional analysis. As studies of the initial variables are continued or completed, new ones are introduced; the experimental design evolves into a series of overlapping individual but integrated longitudinal studies.

The addition of new variables five, ten, or 20 years after the beginning of the original study requires subjects across the entire adult age range. Evolution of the study thus mandates the continuing recruitment of new subjects in the youngest age group as time passes and the population ages. The new subjects in turn provide a built-in opportunity for identification of birth-cohort differences and possible period effects within the population (see below, "Strategies of Analytical Design").

OPERATIONAL CHALLENGES IN LONGITUDINAL STUDIES

Although the longitudinal design is essential to the determination of age changes in individuals, it cannot resolve all the difficulties inherent in cross-sectional studies. Furthermore, longitudinal studies have a number of limitations of their own. What seems to be a simple, straightforward question—"How does aging, or the passage of time, affect performance in individual subjects?"—turns out to be a demon in disguise. Many pitfalls in design, subject selection, data collection, and data analysis may undermine or negate the assumption that changes in serially collected measurements are due to aging. These problems include the following:

1. Recruitment and Screening of Subjects

A primary concern in the selection of subjects for a longitudinal study is their commitment to continued participation in the study and their geographic stability over long periods. This requirement limits the sampling procedures that can be used and must be taken into consideration in the generalization of conclusions drawn from any longitudinal study.

The study may require the inclusion of procedures that prove tedious or distasteful to many people. Subjects selected at random will show both a high initial rate of refusal to participate and a high drop-out rate when presented with a test schedule that includes uncomfortable and time-consuming procedures. To this extent most longitudinal studies, including the BLSA, have compromised true representativeness in order to obtain loyalty and cooperation from their subjects.

Some studies from their inception exclude subjects who present clinical or laboratory evidence of disease. Although this procedure may initially limit the study to healthy subjects, it does not avoid the problem of a subject who subsequently develops chronic illness. If the subject is then dropped from the study a great opportunity to trace the historical development of a disease is lost. Hence subjects who developed diseases were not dropped from the BLSA, although observations made on them were no longer included in analyses of age changes.

Another approach is to accept subjects with diagnosed disease, but to exclude measurements made on them from data analyses designed to characterize normal age changes. Serial observations on these subjects as a subgroup can be of great value in distinguishing the effects of aging from those of aging plus disease.

Although the presence of disease will confound the interpretations about aging in both cross-sectional and longitudinal studies, diseases are more apt to be discovered in subjects in a longitudinal than in a cross-sectional study because of the extended time during which longitudinal subjects are seen and tested. Findings that may be equivocal at one testing can be re-examined on subsequent visits for verification of diagnoses.

2. Attrition

Subject losses must be expected as a longitudinal study progresses. Younger subjects are more likely than old ones to move away from the area of the study or to lose interest and motivation. As the subjects become older, death and disability become major factors (Wilson and Webber, 1976). On the other hand, useful research data may emerge from comparison of measurements in subjects who have survived with those in subjects who have died. This may result in development of new methods of predicting the likelihood of death.

Drop-outs due to loss of contact or to subjects' refusal to continue participation pose a problem in the interpretation of results, particularly when it is evident that those who have left the study differed systematically from those who have remained.

The degree to which findings from longitudinal studies are distorted by attrition, whatever its source, depends on the aspect of aging that is being investigated (see Chapter III). Since some variables are influenced more than others by attrition, each variable in each study must be examined for the drop-out effect.

3. Expansion of Subject Panel

Unlike most longitudinal studies, the BLSA is designed to maintain a specified number of subjects within each age decade throughout its course. When new subjects are introduced, it is important that they resemble the original sample as closely as possible. Ideally, this requires careful description and matching of the original and new populations. Although the BLSA did not attempt such a matching, the self-selection strategy employed in the recruitment of its participants has tended to maintain the character of the sample (see Chapter III).

4. Strategies of Analytical Design

It is often assumed that the differences among serial observations within cohorts followed longitudinally represent the effects of aging. This is not necessarily true: A number of non-maturational effects or factors may also induce differences in serial measurements. Changes in measurements made serially over time may be due to: a) changes in procedures; b) systematic methodological error; c) period effects—environmental or cultural changes that may influence all members of the population under study; or d) aging effects.

An aging effect is present if the dependent variable is a function of age regardless of the subject's birth year or of the period or time of observation. A period effect is present if the value of the variable changes systematically as a function of the time of observation and not as a function of age. A birth-cohort effect is present if the value of the variable changes systematically as a function of the subject's birth year rather than of his age.

Table I.1. A Simple Cross-Sequential Design[a]

Date of Birth	Time of Measurement 1960	 1970
1900	$\bar{X}_{60}$	$\bar{X}_{70}$
1910	$\bar{X}_{50}$	$\bar{X}_{60}$

[a] Subscripts indicate ages

One of the primary difficulties in the analysis of longitudinal data from a single birth cohort is the confounding of period effects with age changes. Traditional longitudinal designs attempt to circumvent the cross-sectional confounding of aging effects with generational or birth-cohort effects by following the same group of individuals over two or more times of measurement, and thus at two or more ages. Such designs, however, are subject to the confounding of age with period effects. Changes that occur between the first and second measurements may be due to intervening historical events rather than to aging; in some tests, previous exposure or practice may be responsible.

Longitudinal changes include age and period effects. Cross-sectional differences include age and cohort effects. Each set of differences is thus influenced by two of the primary effects, those of age, period, and birth cohort. Cross-sequential and time-sequential designs have been proposed to help untangle the confound. In the cross-sequential design, independent samples of individuals from the same birth cohort are compared at different times of measurement, and thus at different ages (Tab. 1). Since a given individual is measured only once, exposure or practice effects are eliminated. In Table 1 the vertical comparison confounds aging and the effects of birth cohort, while the horizontal comparison confounds aging and period effects.

In the time-sequential design, independent samples of individuals of a specified age are compared at different times of measurement (Tab. 2). Age and time of measurement are separated, but both are confounded with birth cohort. No clear-cut statistical separation of age effects from birth-cohort and period effects can be made.

It was originally assumed that, while each of these designs is ambiguous when used alone, it might be possible to separate out age, period, and birth-cohort effects if all were employed and analyzed simultaneously (Schaie, 1965; Baltes, 1968; Riley et al., 1972; Agnello, 1973; and Mason and Mason, 1973). It has since been demonstrated, however, that there is no single solution to the inevitable confounding of the three, and that interpretation of such analysis depends on the data, the goals of the investigator, and the state of knowledge in the area. *Costa and McCrae (1982)* discuss in greater detail the role of judgment in the interpretation of aging, period, and birth-cohort effects.

5. Maintaining Uniformity of Methods and Quality Control

A longitudinal analysis requires special attention to the maintenance of uniformity of tests and testing conditions throughout the study. Continuous quality control is essential. Methods must be examined at regular intervals for consistency of results and stability of standards.

In the BLSA, replicate samples of blood, urine, and tissues are frozen and stored for re-analysis at a later date. Stored plasma samples have made it possible, for example,

Table I.2. A Simple Time-Sequential Design[a]

	Time of Measurement	
Age	1960	1970
60	$\bar{X}_{1900}$	$\bar{X}_{1910}$
70	$\bar{X}_{1890}$	$\bar{X}_{1900}$

[a] Subscripts indicate dates of birth

to validate the methodology used for the determination of cholesterol levels by repeating the analyses at one time on a random subset of samples collected over the entire span of the study *(Hershcopf et al., 1982).*

6. Data Storage and Retrieval

Longitudinal studies generate special problems in the storage and retrieval of data. Thanks to the development of computers, it is now possible to store an immense amount of data in such a fashion that the data can be updated as successive test cycles are completed and at the same time remain available for analysis. It is essential that the system and format of data collection be carefully planned in advance, with the advice of personnel trained in computer technology (Ramm and Gianturco, 1974). It is also essential that special precautions be taken to protect the stored data against catastrophic loss as well as to ensure confidentiality.

7. Staffing

Longitudinal studies pose special problems in the recruitment and maintenance of a research staff. As Busse (1965) has pointed out, scientists who participate successfully in longitudinal studies possess distinctive personal characteristics in addition to their scientific qualifications. They must first of all be patient and willing to wait for longitudinal results to evolve. This does not imply that they will sit with folded hands during the early stages of the study; they will have the insight and initiative to examine data cross-sectionally and to look for significant relations among observations as they accumulate. They will generate new hypotheses that can be explored by the introduction of new tests and procedures. As a result, an effective longitudinal study will be dynamic and will not be limited by the initial test procedures.

Since longitudinal studies are apt to be multidisciplinary in their design, the successful investigator should be able to work with others as a member of a team. Each participating scientist should also be interested in other scientific disciplines and willing to communicate with other scientists in the solution of problems.

8. Financing

Longitudinal studies in adults require stable funding for long periods of time if their full potential is to be realized. Although data analysis must be a continuing part of the program, significant longitudinal results cannot be expected in the early years of a study. Hence, a support system that requires the reporting of substantive longitudinal results at short intervals in order to maintain funding is inappropriate. Since short-term funding has in the past characterized most research support, few individuals or institutions have been prepared to initiate and carry out longitudinal studies.

Of primary concern to research administrators are the presumed high costs and the long-term commitment of resources. However, the ability of a longitudinal study to answer certain important questions about aging answerable by no other technique fully justifies the costs. Although costs may appear high in comparison with those of cross-sectional studies, the potential efficiency of having a population with known characteristics available for multiple satellite short-term cross-sectional studies covering the entire period of adult life greatly increases the cost effectiveness of a longitudinal study. The costs of recruiting multiple groups of well-characterized subjects for short-term studies may well exceed those of maintaining a single stable population.

CHAPTER II

Longitudinal Studies: Past and Present

INTRODUCTION

Although students of growth and development recognized the limitations inherent in cross-sectional studies and the advantages that could be derived from longitudinal studies, the latter were rare before the 1920s. A few classic studies were reported in which height and weight were recorded at frequent intervals in the same children from birth to maturity (Scammon, 1927), but it was not until the 1920s that significant numbers of children were measured repeatedly as they grew and developed (Dearborn et al., 1938; Shuttleworth, 1939; Meredith, 1935). As Chapter I indicates, these early studies provided important insights into growth and development that could not have been derived from cross-sectional observations alone, and thus substantiated the usefulness of longitudinal analyses. Key findings included the discrepancies between growth patterns in individual children and average curves determined cross-sectionally; the estimates of growth rates in individual children; and the relation between growth and specific physiological events, such as maximum growth rates and the initiation of menstruation in girls.

As early as 1947, scientists recognized the necessity of conducting longitudinal studies in adults. The resulting programs (Keys et al., 1961; Dawber et al., 1951) were designed to identify risk factors for the development of cardiovascular diseases (CVD) rather than to describe the phenomenon of aging. Most emphasized physiological functions, and only a few included tests of behavioral or personality characteristics.

During the early 1950s additional longitudinal studies in adults were initiated. In some the goal was to study age changes in specific functions, such as the electroencephalogram (EEG) (Busse and Obrist, 1970) or mental performance (Owens, 1953, 1966; Schaie and Labouvie-Vief, 1974) rather than to identify risk factors for disease. Although these studies focused on aging as a primary variable, only a few made observations on both behavioral and physiological characteristics in the same subjects. Moreover, the number of subjects tested was often small. While most of the studies were purportedly based on "normal" subjects, evidence for "normality" or the absence of specific diseases was seldom adequate. In many studies, self-reports of health status were used: Subjects who said they felt well were regarded as healthy. In studies that assessed health status by a clinical history and physical examination, there were such wide differences in the scope of the examinations that it was difficult to compare the health status of subjects in different studies.

The duration of these studies varied from three to ten years, during which only three or four sets of observations were carried out. None of the studies attempted to calculate age regressions for individual subjects.

Most of the previous studies were terminated before long-term consequences of early events could be evaluated. In some studies observations were limited to subjects aged 65 or older; in others most of the subjects were younger than 25 years at the original testing.

The diversity of populations selected for study, the spectrum of tests, the different testing intervals, and the duration of the various studies were taken into consideration in the design of the Baltimore Longitudinal Study of Aging (BLSA) (see Chapter III). The present chapter, which provides a brief overview of the major longitudinal studies of aging in adults, is intended to supply the background that influenced the design of the BLSA at its initiation and during its subsequent development.

STUDIES OF GROWTH AND DEVELOPMENT

Since about 1920, a number of longitudinal studies have focused primarily on the physical and mental growth of children. Some, such as the Harvard Growth Study (Shuttleworth, 1937, 1939), were concerned primarily with physical growth as it is manifest in anthropometric data. Others that dealt only with tests of intellectual and personality development include the Terman-Stanford study of gifted children (Terman and Oden, 1947, 1959; Oden, 1968). In a few, such as the Denver study (Lewis et al., 1943) and the Oakland Growth Study (Shock, 1946), measurements of a number of physiological functions as well as intellectual, personality, and social characteristics were also made on the same children as they grew and developed. A few, among them the Terman-Stanford study (Terman and Oden, 1947, 1959; Bayley and Oden, 1955; Oden, 1968) as well as the Oakland Growth Study and the Berkeley study (Jones et al., 1971), continued observations into the adult years (Eichorn et al., 1981). Their subjects, many of whom were 50 to 60 years of age in 1978, represent a potential resource for studies of aging if they can be identified and systematically retested.

STUDIES OF SELECTED POPULATIONS OF ADULTS

1. The Cardiovascular Disease (CVD) Project at the University of Minnesota

One of the first studies of adults specifically designed to be longitudinal was initiated in 1947 under the leadership of Dr. Ancel Keys at the University of Minnesota Laboratory of Physiological Hygiene. A total of 281 business and professional men from Minneapolis and St. Paul were recruited for the study, which was "aimed at providing clues about etiology and evaluating the prognostic significance for future heart disease risk of characteristics observed in health" (Keys et al., 1961). The subjects, aged 45 to 54 years in 1947, spent one day each year at the laboratory for an examination which included a review of the interim history, physical examination, nude weight, chest roentgenogram, 12-lead electrocardiogram (ECG), urinalysis, and measurement of hemoglobin and serum cholesterol, as well as special tests that varied from year to year. These included cardiovascular responses to physical exercise, the cold pressor test,[1] cardiovascular responses to passive tilting, ballistocardiogram,[2] EEG, flicker-fusion frequency,[3] basal metabolism, body density, subcutaneous fatness, sugar, uric acid and protein-bound iodine in the blood, anthropometry, and evaluations of personality, based primarily on scores on the Minnesota Multiphasic Personality Inventory (MMPI). Only a part of these longitudinal observations have been published. Leon and her colleagues (1979) reported stability coefficients for the MMPI ranging from .277 (scale 8) to .736 (scale 0) in a sample of 71 men measured at middle age in 1947 and then again 30 years later. In addition, the general profile configuration and two-point code types remained remarkably stable and within normal limits over the 30 years from middle to old age.

---

[1] The rise in blood pressure following the immersion of one arm (or foot and leg) in ice water.

[2] A measurement related to the amount of blood ejected by the heart at each beat.

[3] The frequency of a flashing light perceived as a continuous stimulus.

During the first 15 years of the study (1947–1962), 32 deaths occurred (Keys et al., 1963). Although not all the data have been reported, the study was able to show that elevated serum cholesterol at the first examination and increased systolic blood pressure in the last pre-disease year were significant predictors of the development of coronary artery disease (CAD) and death. Average data over all pre-disease years showed significantly reduced risk among the men in the bottom quartile for diastolic as well as systolic blood pressure, while elevated blood-cholesterol levels were significantly prognostic for the development of CAD.

Keys et al. (1971) reported results in the same group of subjects over a 20-year interval. In 1968, 168 of the 221 living subjects were examined; in 1969, 151 out of 215; in 1970, 153 out of 212. As of 1970, nine of the original 281 subjects had been lost to follow-up, although some of the subjects who were located failed to come to the laboratory to be retested. Information about the health status of the subjects (aged 67–77 in 1970) was obtained periodically from correspondence with the men (49 living outside the state of Minnesota), from their wives, from their personal physicians, or from all three. Of the 60 men who had developed CAD, 42 died or suffered cardiac infarction. Among 20 variables studied, hyper-responsiveness on the cold pressor test, a high level of serum cholesterol, and elevated systolic blood pressure had significant predictive power for cardiac infarction or death from CAD, while smoking and relative body weight (actual weight expressed as a percentage of "standard" weight for sex, age, and height[4]) seemed unimportant as predictors, especially in subjects over age 65.

Analysis of serial measurements of basal metabolism showed that the rate of decrease with advancing age is substantially less than that inferred from cross-sectional studies (Keys et al., 1973).

The availability of longitudinal observations permitted an evaluation of the changes in body weight after cessation of cigarette smoking (Brozek and Keys, 1957). The subjects were men who voluntarily stopped smoking cigarettes; their body weight was measured for two years before and three years after they stopped. A control group of men who continued smoking were matched in age, relative body weight, and actual body weight at the beginning of the first year of the five-year period, without reference to weight trends during the rest of the period. There was no significant difference during the two periods in the body weights of the men who continued to smoke, while those who stopped smoking gained, on the average, 8.2 pounds over the three years following the date they stopped.

There is also a brief report of a 30-year follow-up of these subjects. Obesity, as assessed by the Body Mass Index (wt/ht^2), was not significantly predictive of overall mortality or of death from CAD (Keys, 1980).

---

[4] "Standard weight," based on height, was derived from tables published by the Metropolitan Life Insurance Company (1959).

Table II.1. Duke Study I. Longitudinal Observations: Summary of Variables

- Medical history (original and interim)
- Physical examination
- Neurological examination
- Mental status
- Depression and hypochondriasis
- Dermatological examination
- Ophthalmological examination
  - Visual fields
  - Acuity
  - Color perception
  - Depth perception
  - Color photographs
- Audiometry
  - Pure tone
  - Speech threshold
- Electroencephalogram
- Electrocardiogram
- Chest x-ray
- Laboratory studies
  - Urinalysis
  - Blood morphology
  - Blood chemistry
  - Serologic test for syphilis
  - Cholesterol
  - Urea nitrogen
  - Immunology
- Medical data
- Psychological data
  - Rorschach
  - Aspiration level (TAT)
  - Wechsler Adult Intelligence Scale
  - Reaction time
- Social history and information
- Retirement data
  - Activities
  - Attitudes
- Longevity

From Busse and Maddox (1980)

2. The Duke Studies

Two longitudinal studies have been carried out at Duke University. The goal of the first, initiated in 1955 under the leadership of Dr. E.W. Busse, was to provide answers to two questions: "What are the basic physical, mental, and social processes of normal aging?" and "What accounts for the variations in these processes?" (Busse, 1970). The study was based on 267 men and women, aged 60 to 90 years ($\bar{x}$ age = 70.8 yr), residing in the community. Each panelist was admitted to the Medical Center for a two-day series of medical, psychiatric, psychological, and sociological examinations. These examinations were repeated every two to four years until 1965, and then every two years until 1972. In 1973, the eighth series of tests was completed on the 64 survivors of the original cohort ($\bar{x}$ age = 82.3 yr). In 1976, the 11th series of tests was conducted on 43 survivors ($\bar{x}$ age = 85.2 yr). The tests administered are summarized in Table 1 (Busse and Maddox, 1980). *Normal Aging*, edited by Dr. E. Palmore (1970), reprinted 49 articles by 31 authors based on a variety of analyses of data collected in this panel of subjects.

The results of the first Duke study emphasized the advantages of longitudinal and multidisciplinary studies. The longitudinal analysis made possible the discovery of a general persistence in activities and stability in such traits as hypochondriasis and denial of illness. The inference of failing functions drawn from averages based on cross-sectional data was contradicted by the longitudinal studies, which identified a substantial number of subjects who showed no decline in health status or intellectual function over a number of years; many actually showed improvement in health status (Maddox and Douglass, 1973). Even those who showed substantial impairment of physical functioning, EEG abnormalities, CVD, or impairments in vision and hearing often remained functioning residents of the community, living fairly mobile and independent lives. The surprising degree to which self-perceptions of health vary among individuals constituted a major finding (Maddox and Douglass, 1974).

Table II.2A. Duke Study. II. Adaptation Study Design

	Independent Variables (Resources)		Dependent Variables (Measures of Adaptation)
A.	Physical Medical history Physician's ratings ECG and laboratory tests	A.	Performance Financial and social independence Productivity and social activity Reaction time and accuracy
B.	Psychological Intelligence Personality Vigilance function	B.	Intrapsychic Happiness-depression Psychosomatic symptoms Self-concepts
C.	Social Role and activity measures Socioeconomic status Self-concept Health attitudes Social attitudes		

From Palmore (1974b)

The second longitudinal study, designated "The Duke Adaptation Study," was initiated in 1968. Its goal was to identify the immediate and long-term effects on normal individuals living in the community of such potentially stressful events as death of spouse, serious illness, menopause, children leaving home, preparation for retirement, and retirement, as well as to explore the factors or mechanisms that contribute to "successful" aging defined in a variety of ways (Palmore, 1974a,b). The sample consisted of 261 men and 241 women, aged 45 to 70 years, selected at random from the participants in a major health-insurance plan in the Durham area. The panel was fairly representative of the middle and upper socioeconomic groups in the population. Subjects returned to the center for two days of testing at two-year intervals between 1968 and 1976.

Table 2A shows the design of the Adaptation Study. Table 2B lists the tests administered. Four cycles of testing were completed between 1968 and 1976. Initially, 502 subjects were tested between August 1968 and April 1970. A total of 443 subjects were re-examined by March 1972, 386 were examined for the third time in 1974, and the fourth and final examination was completed in June 1976 on 375 subjects.

In 1974, *Normal Aging II* was published (Palmore, 1974a). Of the 31 articles, 18 are reprints of material previously published in scientific journals, 11 represent new material not previously published, and two are full papers based on presentations made at scientific meetings. Most of the articles deal with results obtained in the first Duke

Table II.2B. Duke Study II. Summary of Variables

Independent Variables (Resources)	Dependent Variables (Measures of Adaptation)
Medical history	Performance
Physical examination	Financial independence
Audiometry	Social independence
Electrocardiogram	Reaction time and accuracy
Chest x-ray	Physical function
Laboratory studies	Intrapsychic
Urinalysis	Life satisfaction
Blood analysis	Happiness
Immunology	Psychosomatic symptoms
Medical summaries	Self-concepts
Psychological data	Mental status
Intelligence	
Personality	
Continuous performance	
Mental status	
Social history and information	
Role and activity	
Socioeconomic status	
Self-concepts	
Health attitudes	
Social attitudes	
Drug-proneness	
Retirement	

From Busse and Maddox (1980)

Longitudinal Study of Aging. Although more details were added from the analysis of additional data, the general conclusions of *Normal Aging* (Palmore, 1970) were not substantially altered. Adult personality showed little change over the eight years of the study.

A more detailed summary of findings of both studies appears in *Final Report: The Duke Longitudinal Studies* and *The Duke Longitudinal Studies on Aging and the Aged* (Busse and Maddox, 1980; 1983).

3. Normative Aging Study

In 1963, the Normative Aging Study began at the Veterans' Administration Outpatient Clinic in Boston under the direction of Drs. B. Bell, C.L. Rose, and Albert Damon. Studies conducted between 1958 and 1963 in a group of 150 ambulatory octogenarian veterans of the Spanish-American War had made it apparent to the investigators that the identification of special characteristics predictive of healthy old age was not possible simply through the study of older individuals but would require serial studies of a group of younger individuals as they aged (Bell et al., 1966, 1972). The goal was to identify the factors that contribute to health in old age by describing the changes that occur with aging.

The subjects, 2032 males, mostly veterans aged 25 to 75 years, living in the Boston area, included representatives of many ethnic, socioeconomic, and occupational groups in the Boston population (Rose, 1965). All were screened for a high level of health at the time they entered the study; subjects with blood pressure higher than 140/90 mm Hg, for example, were excluded. Another selection factor was the

Table II.3. Normative Aging Study (Boston VA Outpatient Clinic)

Tests Administered

- Clinical evaluation
  - Electrocardiogram
  - History and physical exam
  - Chest x-ray
  - Pulmonary function
- Vision
  - Visual acuity
  - Tonometry
  - Dark adaptation
  - Dynamic, static, and flicker perimetry
  - Glare sensitivity
  - Turbidity of the ocular media
  - Depth perception
- Perception
  - Smell
  - Taste
  - Audiometry
- Anthropometry
- Blood chemistry
  - Pepsinogen
  - Triglycerides
  - Protein electrophoresis
  - Plasma testosterone
- Dental evaluation
  - Salivary steroids
- Thyroid function
- Liver function
- Personality tests
- Intellectual function
- Social Information Questionnaire

probability that the subject would remain in the Boston area for his entire lifetime (Rose and Bell, 1965). Although the precise limiting values for exclusion from the study have not been published, the criteria were described as "abnormal values" for pulmonary function, blood-sugar levels, chest x-rays, and ECG. Only four of each ten applicants were accepted after the first clinical evaluation. Subjects came to the laboratory for three nonconsecutive half-days during a five-year cycle. First-cycle tests were administered to approximately 2000 subjects over the period from 1963 to 1968; subsequent cycles were intended to cover the periods from 1969 to 1973 (Cycle II) and from 1974 to 1978 (Cycle III).

The tests included the domains of biochemistry, clinical medicine, oral medicine, neurology, the special senses, anthropometry, psychology, and sociology (Tab. 3). The clinical medicine and biochemistry domains comprised a history, physical examination, standard blood and urine tests, and tests of liver function, serum pepsinogen, triglycerides, plasma-protein electrophoresis, blood sugar, blood-urea nitrogen, uric acid, Ca, P, and protein-bound iodine in the blood. An extensive series of anthropometric measurements was also included, along with resting ECG and blood pressure. ABO blood groupings, lipoprotein phenotyping, and tests for the development of osteoporosis, including a dietary and special medical history and x-rays of the hands, as well as exercise-tolerance ECG, have been added. Smoking histories are obtained for all subjects.

Observations classified as "oral medicine" include orofacial examinations, mastication, a dental survey, facial bone measurements, and parotid saliva secretion rates.

Visual acuity, stereopsis, peripheral retinal shrinkage, dark adaptation, and glare and retinal sensitivity were measured on a subsample of 200 subjects.

Audiologic studies have included sensitivity to frequency and amplitude differences. Tests of retention of verbal and pictorial material, as well as of decision-making under simple and complex classification rules, have been administered to a subsample of the population (Bell et al., 1972).

Selected subjects have been recalled to the laboratory at other times to participate in special tests of such functions as intellectual performance and vision.

Most of the publications from this study represent cross-sectional analyses based on measurements completed during the first cycle of examinations. Burney and Bonus (1972) summarized the clinical laboratory data; Fozard et al. (1972) analyzed the age differences observed in 12 cognitive-performance tests. Dawber and Thomas (1972) presented data originating from clinical examinations of some 1800 subjects, with special emphasis on the frequency distribution of blood-pressure measurements. Clinical data on oral health (Kapur et al., 1972), retinal fields, and pulmonary function (Bell, 1972a) have also been reported for subsamples of the population.

Both Bell (1972b) and Nuttall (1972) presented theoretical approaches to determination of what they called "functional" ages for different types of performance. The approach was to predict chronological age from regression equations derived from specific cross-sectional data sets. For example, Fozard (1972) derived a regression equation to predict chronological age from scores on the General Aptitude test or from the 16 Personality Factor (16 PF) questionnaire. Functional ages were calculated for the domains of blood chemistry, anthropometry, personality, human abilities, sociology, and hearing (Nuttall, 1972) from regression equations derived from experimental data within each domain. The assumption was that meaningful comparisons could be made between the functional ages calculated from observations in different domains. The error of estimate was often large—e.g., 7.2 years for the prediction of age from scores on the General Aptitude test (Fozard, 1972). The usefulness of the concept of "functional age" in contrast to chronological age has been questioned (Costa and McCrae, 1980d).

With completion of the second cycle of testing, analyses were made of the differences in measurements over a five-year interval. Friedlaender et al. (1977) showed that most of the age trends in anthropometry that had been observed in a cross-sectional analysis (Damon et al., 1972) were due not simply to aging but to a combination of aging and birth-cohort effects. *Costa and McCrae (1980a)*, who also analyzed anthropometric findings after a five-year interval, were unable to find evidence of a general aging factor. Age trends, when present, varied so much among individuals that no single trend could be identified.

4. The 1000-Aviator Study, Pensacola

In 1940, a study to determine the value of psychological and physiological tests in predicting success in the flight-training program was initiated at the United States Naval Aviation Center at Pensacola, Florida, by Dr. Ashton Graybiel. An extensive battery of physiological, psychomotor, and psychological tests was administered to selected cadets and officers who entered flight training between July 1940 and May 1941. A total of 1056 subjects, aged 20 to 30 years, were tested. All were preselected, in that they entered the study with supine blood pressure lower than 132/86 mm Hg, and had qualified for flight training by passing rigorous medical and flight-proficiency examinations (Oberman et al., 1965a,b; 1967).

It was not until 1951 that retesting survivors of the original cohort was considered. The study was designed to estimate the current physical status of the men, with particular emphasis on the cardiovascular system, morbidity and mortality rates, and the influence of aviation on these rates. In 1951, 703 of the 829 survivors were re-examined. Survivors were also re-examined in 1957–1958, 1963–1964, 1969–1970,

Table II.4 The 1000-Aviator Study

	Time of Testing					
	1940-41[e]	1951[e]	1957-58[e]	1963-64[e]	1969-70[f]	1977[g]
Mean age	23.6	34.6	41.6	48.6	55.6	63.6
Located	1056[a]	1049	836	811	738	728
Survivors		829	816	794		
Re-examined		703	785	675	675	128
Questionnaires only		115	19	89	43	554
No response		11	12	30		
Died		220[b]	20	17[c]	20	46
Not located		7	3	4		
No contact[d]					61	51

[a] Original sample.
[b] 213 men died in World War II.
[c] 5 men died after returning the questionnaire and were not examined.
[d] No information given on survivorship in the "no-contact" group.
[e] From Oberman et al. (1965a,b)
[f] From MacIntyre et al. (1979)
[g] From MacIntyre (1978)

and 1977 (Oberman et al., 1965a,b; MacIntyre, 1978; MacIntyre et al., 1979). The number of subjects tested at each examination is shown in Table 4. Not all the tests used at the first examination were repeated; Table 5 lists the tests administered at each cycle of the study (Mitchell, 1976).

The blood-pressure and ECG data have been analyzed longitudinally. Although the mean blood pressure for the group showed some increase between the initial examination in 1940–1941 and the re-examination in 1963–1964, blood pressure did not increase with age in every subject; in fact, most of the subjects showed random variations in blood pressure over the first 24 years of follow-up. In this relatively young sample, the average increase was the result of a consistent rise in pressure in a relatively small number of subjects. These, it should be noted, also exhibited greater increments in body weight as they grew older, and had shorter-lived parents (Oberman et al., 1967).

Longitudinal analysis of resting ECG (Harlan et al., 1965) indicated that, of 90 men taken into flight training in 1940 whose ECGs would today be considered as indicating frank or borderline abnormalities, 59 had by 1952 reverted to normal. Of individuals in whom ECG abnormalities persisted, none had developed clinically apparent heart disease at the 1963–1964 examinations.

In 1977, a 37-year follow-up of longevity of the 800 survivors was conducted (MacIntyre et al., 1978). The average age of the group was then 60 years. A mailed questionnaire ascertained the subjects' current health status, presence of significant cardiac problems, current jobs, exercise status, amounts of alcohol and tobacco used, and current weight. A markedly lower death rate than would be expected from a random sample of white American men over a similar period was observed. Lower-than-expected death rates occurred in all three major categories of cause of death: CVD, neoplasms, and accidents.

Table II.5. The 1000-Aviator Study: Summary of Tests Administered[a]

Tests	Time of Testing				
	1940-41	1951	1957-58	1963-64	1969-70
Interview — personal and medical histories	*	*	*	*	*
Physical examination		*	*	*	*
Cardiovascular					
Blood pressure (casual, supine)	*	*	*	*	*
Routine electrocardiogram	*	*	*	*	*
Exercise electrocardiogram			*	*	*
Ballistocardiogram			*[b]	*	*
Vectorcardiogram				*	*
Cold pressor test	*		*[b]		*
Other	*			*	*
Laboratory determinations[c]			*	*	*
Pulmonary and metabolic					
Spirometry	*			*	
Other	*			*	*
Anthropometry					
Somatotype	*			*	
Measurements (in addition to height and weight)				*	*
Teleoroentgenograms		*	*	*	*
Psychologic-psychomotor					
Ataxia test	*			*	*
Tilt chair	*			*	
Other	*			*	
Vision	*			*	*
Neurophysiologic					
Electroencephalogram	*			*	
Audiometry				*	*

[a] Completion of the test is noted by an asterisk (*); if a procedure was not performed during an evaluation, the appropriate column is blank.

[b] Examinations performed on less than 25% of the study group.

[c] Laboratory tests included chest x-ray, lipoproteins, cholesterol, triglycerides, uric acid, glucose, hematocrit, WBC including differential, urinalyses (protein, glucose, microscopic).

From Mitchell (1976)

5. The National Institute of Mental Health Study

In 1955, an extensive multidisciplinary study of aging was initiated at the National Institute of Mental Health under the leadership of Dr. J.E. Birren. The purpose of the study was to examine a broad spectrum of variables in individuals of advanced age in whom disease was absent or minimal. The original focus of the study was on the relations among cerebral physiological changes of advancing age, psychological capacities, and psychiatric symptoms. As the study progressed during its initial five years of operation, social-psychological aspects of old age were added. The resulting study combined the efforts of 22 investigators. The findings of the first study are reported in detail in *Human Aging* (Birren et al., 1963).

The subjects were 47 male volunteers aged from 65 to 91 years (median age = 71 yr) who were living in the community. All subjects were reported to be healthy on the basis of a detailed medical examination. Subjects were admitted to the Clinical Center of the National Institutes of Health for a period of two weeks, during which an extensive battery of physiological and psychological tests was administered (Granick and Patterson, 1971).

In 1961, 29 of the 39 survivors were re-examined at the Clinical Center, Bethesda. Although no comprehensive report of the follow-up has been published, some aspects of the work were reported by Butler (1967), Botwinick and Birren (1965), and Birren (1964). The interviews and tests administered are shown in Table 6. The general

Table II.6. The National Institute of Mental Health Study: Interviews and Tests Administered

Examination or Test	Period of Study		
	1956	1961	1967
Medicine and physiology			
Medical history	•	•	•
Physical examination with complete neurological	•	•	•
Hematology	•	•	•
Blood chemistry	•	•	•
Urinalysis	•	•	•
Chest x-ray	•	•	•
Skull x-ray	•	•	•
Electrocardiogram	•	•	•
Electroencephalogram	•	•	•
Pulmonary-function studies	•	•	•
Cerebral-blood-flow studies	•	•	•
Audiometric examination	•		•
Click-perception tests	•		
Delayed auditory feedback tests	•		
Psychological			
Addition rate	•	•	•
Arithmetic alternation rate	•	•	•
Draw-a-Person	•	•	•
Emotional projection test	•		
Family scene	•		•
Homonyms	•	•	
Learning	•		
Level of aspiration	•	•	
Minnesota Multiphasic Personality Inventory	•		
Mirror tracing	•		
Perception of line diffrence	•	•	
Raven Progressive Matrices	•	•	•
Reaction time	•	•	
Rorschach	•	•	
Sentence-completion test	•		•
Speed of card sorting	•	•	
Speed of copying digits	•	•	•
Speed of copying words	•		•
Stroop Test	•		•
Thematic Apperception Test	•		
Wechsler Adult Intelligence Scale	•	•	•
Weigl Color Sorting	•	•	
Wisconsin Card Sorting	•	•	
Word fluency	•		

Table II.6. The National Institute of Mental Health Study: Interviews and Tests Administered—*(Cont'd.)*

Examination or Test	Period of Study 1956	1961	1967
Social-psychological interview			
Family history (or interval history)	•		•
Educational history	•		
Occupational history	•		
Retirement planning and activities	•	•	•
Marital history	•	•	•
Living arrangements	•	•	•
Use of time	•	•	•
Social relations and interaction	•	•	•
Attitudes toward life	•	•	•
Goals and aspirations	•	•	•
Critical turning points in life	•	•	•
Significant losses	•	•	•
Observed physical and mental changes in aging	•	•	•
Psychiatric interviews			
History of psychiatric contact	•	•	•
Personal-social history (or interval history)	•	•	•
Psychiatric-symptom check list	•	•	•
Mental-status evaluation	•		•
Assessment of attitudes about			
futurity	•	•	•
death	•	•	•
self	•	•	•
aging	•	•	•

From Granick and Patterson (1971)

results indicated little change in the subjects tested after the five-year interval (Butler, 1967), except in tasks such as card-sorting (Botwinick and Birren, 1965) in which speed of performance was the criterion.

In 1967–1968, 19 survivors from the original group of 47 were tested for the third time by other investigators at the Philadelphia Geriatric Center; findings were reported in *Human Aging II* (Granick and Patterson, 1971). Twenty-four of the subjects had died, and four had dropped out. Not all subjects participated in all the tests; for example, cerebral blood flow was measured in only eight subjects.

Every attempt was made to replicate the methods used in the first testing, and many of the investigators who had taken part in the first testing participated in the analysis of the data. Longitudinal results are summarized in Table 7 (Granick and Patterson, 1971). Table 8 lists the variables that at the initial testing had shown differences between subjects who survived and non-survivors.

Although the number of subjects was small, this study showed that decrements in intellectual performance with advancing age were significantly greater in subjects who also developed CVD than in subjects who remained healthy.

6. The Basel, Switzerland, Study

The Basel study, organized by Drs. F. Verzar and O.R. Gsell, was based on measurements made at one- or two-year intervals over a ten-year period (1955-1965) in 121 male subjects aged from 8 to 85 years. Most of the subjects were aged 26 to 56 years at the beginning of the study (1955) and were employees of the CIBA

Table II.7. The National Institute of Mental Health Study: Changes in Survivors after 11 years

Study Area	Type and Direction of Change
Medicine	Almost half had significant diseases. Erythrocyte sedimentation rate increased in healthy subjects.
Cerebral physiology	
Circulation	Cerebral blood flow decreased.
Electroencephalogram	8 subjects showed some type of EEG change. No systematic changes in the group, but 5 subjects showed slowing of the dominant occipital rhythm. Changes found in all 4 cases of chronic brain syndrome.
Psychology	Vocabulary improved. Picture arrangement improved. Speed of addition declined. Speed of arithmetic alternation declined. Speed of copying words declined. Quality of Draw-a-Person declined. Quality of sentence completions declined.
Psychiatry	Trend toward more organic mental changes. Self-monitoring of intellectual and physical capacities more prominent. "Energy" decreased (self-reports). Sexual interest declined.
Social psychology	Losses in social environment increased. Vulnerability to failure at coping with stressful events increased. "Energy" decreased (self-reports).

From Granick and Patterson (1971)

Pharmaceutical Company (Verzar, 1967; Tripod, 1967). The measurements included anthropometry—height, weight, and circumferences of abdomen, throat, and wrist—vital capacity, maximum expiratory volume, ECG, systolic and diastolic blood pressure, pulse-wave velocity, range of accommodation of the eye, and state of health (Gsell, 1967, 1973). Longitudinal analysis of the data was accomplished by a) calculation of average values obtained from this group of subjects as they aged; b) presentation of curves based on serial measurements on the same subject; and c) calculation of average changes over the ten-year interval. The number of subjects on whom data were available for the full ten years varied from 72 (health evaluation, anthropometry, blood pressure, ECG) to 17 (vital capacity).

Brückner (1967) found that the longitudinal analysis of individual changes with aging in the range of accommodation of the eye showed a more rapid decrease with age than that predicted from cross-sectional studies. Similar differences between cross-sectional and longitudinal analyses were reported by Monnier (1967) for the age-related increase in pulse-wave velocity, which was attributed to the development of arteriosclerosis in major blood vessels. The increase with age shown in 27 subjects for whom serial observations were available over the ten-year period was greater than that shown by the average cross-sectional curve.

Table II.8. The National Institute of Mental Health Study: Significant Differences between Survivors and Nonsurvivors (Initial Measurements)

Study	Factor	Direction for Survivors
Medicine	Health status (groups I and II)	Healthier
	Systolic blood pressure	Lower
	Diastolic blood pressure	Lower
	Mean arterial blood pressure	Lower
	Weight	Heavier
	Arteriosclerosis	Less
	Chronic cigarette smoking	Less
	Serum cholesterol in those who died from coronary heart disease	Lower
	Serum albumin in those who died from carcinoma	Higher
Cerebral physiology		
Circulation	None	
Electroencephalogram	None, but a tendency shown with respect to:	
	Peak occipital frequency	Higher
	Percentage fast activity	Lower
Psychology		
Intellectual and psychomotor	WAIS Verbal scale	Higher
	WAIS Performance scale	Higher
	WAIS subtests:	
	Vocabulary	Higher
	Information	Higher
	Comprehension	Higher
	Similarities	Higher
	Digit symbol substitution	Higher
	Block design	Higher
	Speed of copying digits	Higher
	Speed of copying words	Higher
	Principal Component I (stored information)	Higher
Personality	Draw-a-Person	Higher
	Rorschach	Higher
	Homonyms	Higher
	MMPI-Si scale (social involvement)	Higher
Psychiatry	Adaptation	Better
	Mental status	Higher
Social psychology	Organization of behavior	Higher
	Environment loss	Lower

From Granick and Patterson (1971)

The Basel study provides evidence that the following age differences, identified in cross-sectional studies, represent age changes that occur in individuals (Gsell, 1967): rise in blood pressure, increase in obesity, decrease in vital capacity, decrease in range of accommodation of the eye, increase of pulse-wave velocity, and increase of abdominal and chest circumference.

Table II.9. The Bonn Longitudinal Study of Aging:
Number of Participants and Dropouts and Number of Control-Group Participants

	Measurement Point					Control
	I	II	III	IV	V	
Younger cohort						
Men	59 (4)[a]	55 (8)	47 (7)	40 (4)	36	13
Women	55 (5)	50 (2)	48 (10)	38 (7)	31	22
Sum	114 (9)	105 (10)	95 (17)	78 (11)	67	35
Older cohort						
Men	59 (7)	52 (3)	49 (10)	39 (11)	28	13
Women	49 (4)	45 (5)	40 (11)	29 (3)	26	13
Sum	108 (11)	97 (8)	89 (21)	68 (14)	54	26
Total	222 (20)	202 (18)	184 (38)	146 (25)	121	61

[a] Numbers of dropouts are indicated in parentheses.
From Rudinger and Schmitz-Scherzer (1976)

7. The Bonn, West Germany, Study

The Bonn Longitudinal Study of Aging was initiated by Dr. H. Thomae in 1965. The sample consisted of 220 men and women, born between 1890 and 1905, from different parts of West Germany. Each subject was examined at five different times during the period between 1965–1966 and 1972–1973 (Thomae, 1976). A control group of 61 new subjects was examined in 1972–1973. Table 9 shows the number of subjects tested at each period.

Since the goal of the study was to explore the factors involved in the marked variation among individuals apparent from cross-sectional studies of aging, a global approach, which relied heavily on interviews, was taken. At each measurement point, one three-hour interview was devoted to the assessment of present situation, one interview of two to four hours to that of the past (childhood, adolescence, young adulthood), and a third to assessment of the future outlook. Each interview was tape-recorded so that future analyses could be made in the light of new developments in theories about the psychological and social aspects of aging. In addition, standard psychological (WAIS; Raven Progressive Matrices), personality (Thematic Apperception Test; Rorschach; Riegel Scales), and psychomotor tests (choice and simple reaction time), and a brief physical examination to estimate health status were administered. The interviews focused on social conditions, social involvement, leisure activities, education, occupation, housing, life history, and outlook on the future.

The Bonn study, placing major emphasis on psychological and personality characteristics, highlights the marked individual variations in aging. More than 65 publications dealing with specific aspects of the data have already appeared (Thomae, 1976). The overall findings of the study point to the importance of interactions among health status, psychological competence, and social and economic conditions in determining patterns of aging. The richness of individual compensatory adaptations indicates that there are multiple pathways toward successful aging—a conclusion that can be drawn only from longitudinal observations.

STUDIES OF COMMUNITY POPULATIONS

The longitudinal studies described above have been conducted in relatively small selected populations. Other studies have used larger numbers of subjects selected from the total population of a community. The Framingham, Tecumseh, and Atomic Bomb Casualty Commission studies are examples.

1. The Framingham Study

In 1948, a prospective study designed to identify the relations of age, sex, family history, occupation, educational level, national origin, serum-lipid levels, smoking history, and physical activity to the development of CAD was initiated under the sponsorship of the United States Public Health Service. On July 1, 1949, the program was transferred to the newly established National Heart Institute.

The primary goal of the study was "the determination of factors influencing the development of heart disease." In order to meet this goal, it was recognized that repeated examinations of a large number of subjects from a community would be required. For this purpose two thirds of the 30- to 59-year-old population of Framingham, Massachusetts, were selected from published lists of all residents of Framingham over the age of 20 (Dawber et al., 1951). The process yielded 6507 individuals who were invited to participate in the study. Of this group, 4469, or 68.7% of the drawn sample, came to the study clinic for the initial examination, which was accomplished between 1948 and 1952. An additional group of 740 "volunteers" who were not selected by the formal sampling procedure were added to the original respondents. The total number of men and women tested at first visit was thus 5209 (Dawber et al., 1951; Gordon et al., 1959; Dawber et al., 1963; The Framingham Study, 1968–1974).

The study was designed to re-examine each subject every two years over a period of 20 years. At the eighth examination, 14 years after the first, 4678 subjects (89.8% of the original population) were alive; of these, 4030 (86.1% of those still alive and 77.4% of the original group) reported for examination.

By 1977, 3680 subjects were still alive and 1529 had died. Although the two-year test interval has not been strictly maintained since 1970, certain observations are continuing, with special emphasis on the factors involved in the development of cerebrovascular disease. The measured variables of major interest as risk factors for the development of CVD and for death are listed in Table 10. The data were analyzed longitudinally at the end of ten and 20 years, and numerous publications have appeared (The Framingham Study, 1974–1978).

The outstanding achievements of the study include the identification of significant risk factors for the development of CVD and the demonstration that they may be additive or multiplicative in their effects. The risk factors or predictors that have been identified are: cigarette smoking, elevated blood pressure, elevated serum cholesterol and low-density lipoproteins, low vital capacity, diabetes, and obesity. In addition, certain ECG abnormalities and x-ray evidence of cardiac enlargement are predictive (Kannel, 1978). These findings, which are convincing because of the longitudinal or prospective design of the study, have had a significant impact on public-health programs to reduce the incidence of CVD.

Table II.10. The Framingham Study

Major Independent Variables (Risk Factors)	Major Dependent Variables (End-Points)
Physical examination	Mortality
Blood pressure	Total
Height and weight	Coronary heart disease (CHD)
	Sudden death from CHD
History	Non-sudden death from CHD
Alcohol consumed	Cardiovascular, non-coronary
Cigarettes per day	Non-cardiovascular
Laboratory tests	Morbid events, vascular
Blood or serum	Heart
Glucose	CHD
Cholesterol	Myocardial infarction
Phospholipids	Angina pectoris
Hematocrit and hemoglobin	Congestive heart failure
Uric acid	Brain
Urinalysis	Cerebrovascular accident
Chest x-ray	Brain infarction
Vital capacity	Peripheral arterial
Electrocardiogram	Intermittent claudication

Compiled from data in The Framingham Study (1978)

2. The Tecumseh, Michigan, Community Health Study

The Tecumseh study, first planned in 1957, was designed as an ecologic investigation of an entire community, in contrast to studies based only on samples. The study was designed to identify the early origins of impaired health in order to detect precursors of overt illness at a time when preventive measures might be instituted (Epstein, 1960; Napier, 1962; Epstein et al., 1965; Montoye et al., 1965; Napier et al., 1970; Montoye, 1975; Montoye et al., 1978).

Although all residents of the town and its environs were contacted, approximately 80% or 8641 (4239 men, 4402 women) were actually tested between 1959 and 1960. The subjects ranged in age from birth to 70+ years (the 70+ group consisted of 127 men and 162 women, 3% of the total group). Approximately 45% of the subjects were younger than 20 years at the first testing.

In the years 1961 to 1965, the second cycle of examinations was conducted. A total of 9226 subjects or 82% of the population, including 2499 new residents, was tested.

A third cycle of examinations was conducted from February 1967 through June 1969. This cycle was limited to persons who had been examined at least once earlier and persons of all ages, whether previously examined or not, who resided in a 10% random sample of dwelling units. A number of relatives of subjects identified as having CAD or diabetes mellitus were also tested. Subjects tested in each of the three cycles totaled 4312.

Table 11 lists the tests administered in each cycle. Reports of cross-sectional analyses by age decades have been published for physical-activity levels, heart-rate response to exercise, oxygen uptake, ECG, proteinuria, and ventilatory response to exercise (Montoye, 1975); glucose tolerance (Montoye et al., 1977); forced vital capacity and other pulmonary-function tests (Higgins and Keller, 1973); and serum cholesterol and disease prevalence (Epstein et al., 1965). The testing program was terminated in 1970.

Table II.11. The Tecumseh Study: Tests Administered

	Cycle 1 1959–1960	Cycle 2 1961–1965	Cycle 3 1967–1969
Subjects			
Men	4239	4479	2857
Women	4402	4747	3155
Total	8641	9226	6012
New subjects		2499	
% of Tecumseh population tested	80	82	
History and physical exam	•	•	•
Anthropometry			
Height and weight	•	•	•
Skinfolds	•	•	•
Skeletal diameters	•	•	•
Laboratory tests			
Electrocardiogram			
Resting	•	•	•
Exercise step test		•	•
Treadmill exercise			•
Pulmonary function			
Vital capacity	•	•	•
FEV$^{1.0}$	•	•	•
Nitrogen washout			•
Roentogenographic studies			
Chest	•	•	•
Hand		•	
Cervical spine		•	
Fasting and one-hour glucose-tolerance test	•	•	•
Blood tests			
Hemoglobin	•	•	•
Blood type (11 groups)			•
Uric acid	•	•	•
Cholesterol	•	•	•
Triglycerides			•
Lipoprotein electrophoresis			•
Rheumatoid factors	•	•	•
Urinalysis	•	•	•
Skin sensitivity to tuberculin and histoplasmin (20% of population)		•	•
Activity questionnaire		•	•

From Montoye (1975)

Publications from the study have analyzed data by the cross-sectional method, with special emphasis on the role of physical activity in the maintenance of health. Longitudinal analyses of repeated observations on the same subjects have not been published.

3. The Atomic Bomb Casualty Commission Study

Almost immediately after the explosion of the atomic bombs over Hiroshima and Nagasaki and termination of hostilities, steps were taken by United States military

Table II.12. Atomic Bomb Casualty Commission—Hiroshima Study: Tests Administered to Clinical Subsample

On All Subjects at 2-Year Intervals

- Height and weight
- Skinfold thickness
- Hematology
  - Hemaglobin, hematocrit, red-cell count, white-cell count, differential white count, sedimentation rate
- Urinalysis, including microscopic examination
- Stool examination for blood, ova, and parasites
- Blood pressure
- Chest x-ray
- Questionnaire—history and present symptoms
- Interview information on smoking, drinking, diet, occupation, and residential facts

On Part of the Sample at Varying Times

- Chromosome studies (number and aberrations) in culture of lymphocytes
- Cardiovascular disease
- Grip strength
- Anthropometric measurements
- Vital capacity
- 12-lead electrocardiogram
- Serum cholesterol, triglycerides, and uric acid
- Blood clotting and lysis
- Ocular fundus photography

Compiled from data in Hollingsworth et al. (1965), Finch and Beebe (1975), Belsky et al. (1973)

authorities to assess the immediate and long-term effects of exposure to radiation on longevity and the incidence of disease, especially cancer, among humans. Over a period of years the planning and supervision of the studies were transferred to the Atomic Bomb Casualty Commission (ABCC), whose members were appointed by the National Academy of Sciences of the United States and the Japanese National Institute of Health.

The primary focus of the studies was epidemiological, a comparison of morbidity and mortality from a variety of diseases in survivors of the explosions with morbidity and mortality in residents of Hiroshima and Nagasaki who were not exposed to radiation.

It was also recognized that steps should be taken to observe the development of any abnormalities induced by radiation exposure through repeated observations in the same subject over time—the longitudinal approach. Many radiobiologists at this time (about 1948–1952) were of the opinion that exposure to radiation simply "accelerated normal aging;" to test the hypothesis, serial observations in both the exposed and the non-exposed were necessary.

In 1958 a sample of about 20,000 subjects was drawn by statistical techniques from a total population of 110,000 identified as residents of Hiroshima at the time the bomb was dropped. The 20,000 were about equally divided among four groups of about 5000 each who had been within 2000, 2000 to 2500, and 2500 to 10,000 meters of the epicenter, or had not been exposed to radiation because they had been out of the city when the bomb was dropped. Observations on the subjects in the non-exposed group provided control data.

The design of the study called for detailed medical examinations, with some laboratory tests to be conducted on each of the 20,000 subjects at two-year intervals. Although the task proved greater than the resources, a great many serial examinations were completed. Since the primary interest of the study was to identify the long-term effects of exposure to radiation, analysis of age effects was not given a high priority. In addition to a standard medical history and physical examination, height, weight, blood pressure, chest x-ray, blood morphology, hemoglobin, and routine urine examinations were made in practically all subjects (Tab. 12). Additional tests, including extensive anthropometry, skinfold thickness, grip strength, vital capacity, 12-lead ECG, serum cholesterol, triglycerides, and uric-acid levels in the blood were conducted in subpopulations over varying periods of time. Cross-sectional analyses were used, but no publications have presented results based on longitudinal analyses.

In April 1975 responsibility for all studies was transferred to a newly established Japanese Radiation Effects Research Foundation, under which studies of the long-term effects of exposure to ionizing radiation are being continued (Atomic Bomb Casualty Commission, 1978).

Most of the publications from the ABCC study are concerned with determining the effect of exposure to various levels of radiation on the incidence of disease, especially cancer. In general, exposure to non-lethal doses of radiation did not seem to have any significant long-term effects. Radiation injuries were primarily confined to the period immediately following acute exposure, and the incidence of symptoms (epilation, bleeding, and oropharyngeal lesions) increased almost linearly from 5%–10% among those exposed to a total dose of 50 rad to 50%–80% among those exposed to about 300 rad, beyond which the proportion leveled off (Atomic Bomb Casualty Commission, 1978).

Hollingsworth et al. (1965) attempted to devise an index of physiologic age based on measurements of a number of physiological variables in 437 non-exposed males and females, aged from 10 to 70 years, from the ABCC Adult Health Study. Scores on nine tests—skin elasticity, systolic blood pressure, pulmonary vital capacity, hand-grip strength, visual reaction time, vibratory perception at the ankle (120/sec), visual acuity (Snellen), auditory threshold (decibels at 4000 cycles/sec), and serum cholesterol—were combined in a multiple regression equation to provide a "physiological age" score for each individual. When a value for "physiological age" was derived, it was highly correlated with chronological age. The observations, however, were not analyzed to determine whether subjects whose physiologic age was lower than their chronological age had lived longer. Such a test for the validity of an index of physiological age is essential; but until follow-up observations have been made over a period during which a substantial number of the subjects have died, no definitive answer can be given.

Examinations of mortality data on a large sample of Japanese over the period from 1950 to 1972 provided no support for the hypothesis that exposure to ionizing radiation accelerates aging. The effects appear to be the result of specific radiation-induced diseases, especially neoplasms (Finch and Beebe, 1975).

Dock and Fukushima (1978) analyzed observations of blood pressure in 13,814 Japanese subjects who were participants in the ABCC Adult Health Study. Blood pressures were measured at two-year intervals between 1958 and 1972. At the seventh examination, in 1970–1972, 8707 or 72% of the 12,123 subjects still in the study were examined. In subjects younger than 60 years, there was a small but consistent rise in average values for both systolic and diastolic blood pressure with increasing age over the period. Subjects in a recent cycle exhibited pressures about the same as or slightly lower than those of subjects who had been the same age a decade or more earlier. Subjects aged 60+ at the time of the initial test failed to show any further increase in pressure during the subsequent years. Very few "mild" hypertensives progressed into the "severe" category during the period of study. Initial pressure and age were the primary predictors for the development of hypertension. Subjects who had shown a transient rise in blood pressure at their first visit had a higher risk of an increase in blood pressure over the following decade than subjects with stable values during the first visit.

STUDIES OF SPECIFIC VARIABLES

1. Intellectual Functions

A number of studies have used the longitudinal method to investigate age changes in specific domains. Owens (1966) reported on the serial analysis of Army Alpha test scores for intelligence of 96 men tested in 1961 ($\bar{x}$ age = 61 yr) who had originally been tested as entering freshmen at Iowa State University in 1919. They represented survivors of a group of 127 subjects who had been retested in 1949–1950 at a mean age of 50 (Owens, 1953). Of the 31 subjects lost between 1949–1950 and 1961, 13 had died, five were disabled, five could not be located, and eight refused to participate. The important finding of the study was that within individuals there was little change in test scores over the ten-year interval from 1950 to 1960. The previous report (Owens, 1953) had shown increments in verbal ability and total score over the interval between 1919 and 1950 (age span 20–50 yr), although numerical ability showed a slight decrement. This study was the first to raise serious doubts about the presumed decline in mental abilities with advancing age that had been inferred from previous cross-sectional studies (Jones and Conrad, 1933).

In 1921–1922, a group of 1528 intellectually gifted children, whose age range was from three to 19 years, was selected by means of intelligence tests. The study, organized by L.M. Terman, is known as the Terman-Stanford study of gifted children. The subjects ($\bar{x}$ I.Q. = 151, range 140–200) have been followed since 1921 by means of field-worker interviews, questionnaires, tests, and personal correspondence. Special tests of intellectual function were administered to 954 subjects ($\bar{x}$ age = 29.5 yr) in 1939–1940, and to 768 subjects ($\bar{x}$ age = 41.5 yr) in 1952. An analysis by Bayley and Oden (1955) of changes in test scores over the 12-year interval showed that, in general, scores on the mental tests increased in individual subjects between the ages of 30 and 40 years.

Schaie and his co-workers have examined the changes in intellectual performance with age. The five subtests of the SRA Primary Mental Abilities (Thurstone and Thurstone, 1949) and Schaie's test of Behavioral Rigidity were administered to subjects drawn on three occasions (1956, 1963, and 1970) from approximately 18,000 members of a prepaid medical plan operating in the Pacific Northwest. In 1956, 25 men and 25 women in each five-year interval from 21 to 70 years of age were tested. In 1963, 302 of the original 500 subjects were retested; 161 were tested for a third time in 1970. In addition, 960 subjects were tested for the first time in 1963; of these, 409 were retested in 1970. Another 701 men and women were tested for the first time in 1970. As a result, it was possible to measure seven- and 14-year longitudinal changes, and to estimate mean changes by comparing independent samples from the same birth cohorts at the seven- and 14-year intervals. Schaie and Labouvie-Vief (1974) concluded that much of the apparent age decline was in reality due to birth-cohort differences and that aging did not result in an inevitable decline in all intellectual functions.

2. Physiological Variables

Blood pressure. Jenss (1934) examined in detail the rate of change of systolic blood pressure with age in individual male subjects. The data were obtained from annual physical examinations of 1139 United States Army officers (age range = 27-54 yr) between 1916 and 1930. Each subject provided at least seven observations ($\bar{x} = 9.4$ observations over $\bar{x}$ period of 12.3 yr, range = 7-14 yr). The longitudinal analysis showed that age differences in systolic blood pressure derived from cross-sectional data were not representative of all the officers. Some individuals showed an appreciable decline in pressure, others a corresponding rise, during the period of observation. Regression toward the mean was observed: Officers with an initial reading considerably above the average tended to show a decrease in systolic pressure with age, those with a low initial pressure an increase. The inverse relation between the initial level and the trend of systolic pressure was more marked in the young than in the old officers.

Engel and Malmstrom (1967) analyzed blood-pressure measurements made annually over a period of 20 years on 242 male employees of the Travelers Insurance Company. Although a single linear regression of blood pressure on age could be fit to the observations on most of the subjects, 56 were identified in whom two straight lines of different slopes were needed to describe the data. The ability to detect a change in slope or any form of non-linearity obviously requires multiple longitudinal observations over a substantial number of years.

Exercise. In a study by Robinson et al. (1975), the physiological adjustments to both aerobic (5.6 km/hr, at 9% grade) and maximal treadmill work, which had first been determined when the subjects were college students, aged from 18 to 22 years, were compared with measurements made in 20 of the same subjects at ages 40 to 44 and again at ages 49 to 53. In the aerobic walk, oxygen consumption increased in proportion to the gain in body weight, but efficiency did not change with age; the amount of oxygen consumed per unit of work remained the same. The men who had gained the most weight showed the greatest elevations of heart rate and blood lactate after exercise, with a significant decrease in the efficiency of lung ventilation.

The most pronounced changes with age occurred in metabolic, respiratory, and circulatory adjustments to maximal or exhausting work. Over the 30 years of the study, maximum oxygen consumption declined between 20% and 30% in individual subjects. The most important finding, however, was that eight of the men showed an average increase of 11% in maximum oxygen uptake between the ages of 40–44 and 49–53 years. This increase in physical fitness was associated with increased participation in vigorous activities and cessation of smoking.

Dill et al. (1967) assessed the physiological state of 16 runners on two occasions over a span of 20 years. Many of the subjects had been Olympic contenders, a few world champions; they displayed unusual physical fitness at the time of their first testing. In addition to a clinical assessment including chest x-rays, ECG, urinalysis, and measurement of serum cholesterol, pulmonary-function tests were carried out. Observations made with the subjects in the basal state included oxygen consumption, CO_2 production, respiratory rate and minute volume, heart rate, and blood pressure. Arterial blood samples were analyzed for hemoglobin oxygen saturation, pCO_2, pO_2, and CO_2 combining power. Physiological responses to grades of exercise performed on a treadmill were measured. Although complete observations were not obtained on all subjects, age trends for the group as well as for the individual subjects could be identified. The major finding was that age differences based on average values for the group did not predict age changes found in individual subjects tested after an interval of 20 years. Although maximum oxygen uptake decreased in all the subjects, the amount of the decrease varied widely among subjects. Despite the absence of statistically significant results, the authors concluded that regular strenuous exercise and the absence of cigarette smoking seemed to be conducive to a high level of physical fitness.

Asmussen et al. (1975) re-examined 19 men ($\bar{x}$ age = 60.7 yr) and six women ($\bar{x}$ age = 63.2 yr) who had been tested first as students of physical education ($\bar{x}$ age: men, 23.9; women, 23.5 yr) and later at mean ages 48.7 (men) and 51.3 years (women). Measurements made under resting rather than basal conditions included blood pressure, heart rate, O_2 uptake and CO_2 production, vital capacity, residual air, handgrip strength, and reaction time to light and sound. Exercise testing (Krogh bicycle ergometer) at 617 kpm/min (men) and 411 kpm/min (women) was used to determine O_2 uptake. Average curves showed significant age decrements in vital capacity, handgrip strength, speed of reaction, and maximum oxygen uptake during exercise. Systolic blood pressure increased with age. Decrements seemed to be greater between ages 48 and 62 than between 23 and 48. The authors concluded that the changes in physiological functions they had observed over a period of 40 years (at ages 20–60) seemed to be identical with the results of most cross-sectional studies of the same functions.

Åstrand et al. (1973) retested 35 female and 31 male physical-education students (aged 20–33 yr at initial testing) after an interval of 20 years. The subjects performed submaximal and maximal exercise on a bicycle ergometer. Although at the second testing maximum oxygen uptake had declined by approximately 20%, wide individual differences were found. Average values for maximum oxygen uptake were, however, in good agreement with the findings of other cross-sectional and longitudinal studies.

Dehn and Bruce (1972) measured maximal oxygen uptake during a multistage treadmill test in 86 healthy men between 40 and 72 years of age. In 40 of the subjects similar tests had been made two to three years earlier. A cross-sectional analysis revealed an average annual decrement in maximum oxygen uptake of 0.28 ml/min·kg. Combined data from 17 cross-sectional studies found in the literature showed an average annual decrement of 0.40 ml/min·kg. Longitudinal analysis of the repeated observations on 40 subjects of the present study revealed a significantly greater annual decrement of 0.94 ml/min·kg, which compared remarkably well with longitudinal results previously reported by Dill et al. (1967).

Although both cross-sectional and longitudinal analysis indicate that on the average maximum oxygen uptake tends to fall with age, longitudinal studies show that the downward trend may be reversed in middle-aged subjects. It is assumed that such changes in life style as elimination of cigarette smoking and the introduction of regular regimens of exercise contribute to this reversal.

SUMMARY

This brief review of the major longitudinal studies that have been or are being conducted on adults illustrates the great diversity in design, subject selection, and tests administered. This diversity is a source of both strength and weakness. The strength lies in the broad spectrum of tests that have been used. The weakness is the limited possibility for comparisons between different types of performance within the same subject. Few studies have included a wide range of both psychological and physiological tests in the same subjects.

Table 13, which summarizes the major features of selected longitudinal studies of adults, shows the diversity in their design and execution and emphasizes the fact that the ideal longitudinal study is yet to be designed.

It is inevitable that many serial observations that have been collected have never been analyzed longitudinally with respect to aging. This omission seems often to have been due to the emergence of unanticipated methodologic complexities in the longitudinal analysis of data; not uncommonly, it has resulted from loss of critical personnel or fiscal resources. Although several of the studies enumerated in Table 13 had never had as their primary goal an assessment of the age changes for their own sake, their data bases offer the potential for such analysis.

The criteria for subject selection have varied widely. In most studies the number of subjects is relatively small; groups are often highly selected and usually male. Women are grossly underrepresented, and no studies have included blacks or other ethnic groups as identifiable subsets.

Since longitudinal studies have been initiated by groups of investigators with different interests and goals, communication among investigators has often been minimal. In order to improve communication among investigators in longitudinal studies, a series of six conferences held between 1972 and 1975 (Rose, 1976) brought together representatives from eight longitudinal studies: Baltimore Longitudinal Study of Aging; Normative Aging Study; The Framingham Study; the Duke University Studies of Human Aging; the 1000-Aviator Study; the Tecumseh, Michigan, Community Health Study; the CVD Project at the University of Minnesota; and the Kaiser Permanente Foundation Study, Oakland, California (not reviewed here because of the absence of documentation.)

The meetings provided a forum for the interchange of information and discussion of problems associated with the operation of longitudinal studies. The question of pooling data for specific tests from a number of studies was explored. Data for height, weight, blood pressure, cholesterol, and vital capacity obtained from five longitudinal studies (Duke, Tecumseh, Baltimore Longitudinal Study of Aging, Minnesota, and Normative Aging Study) were compared (Garvey, 1973). The results showed that pooling of original observations from different studies was impractical because of systematic differences in the selection of population and in methods of measurement. Although direct pooling of data may not be feasible, the results of longitudinal analyses from different studies can be compared. Age trends found in a number of studies are apt to support more valid generalizations than those found only in a single study. On the other hand, results found in some populations but not in others may provide important epidemiological clues to demographic, genetic, or environmental variables that influence the responses to aging.

Table II.13. Selected Longitudinal Studies of Adults

Study	Began	Ended	Nature of Sample	N	Sex & Ages at Entry	Test Interval	Test Period	N Repeat Cycles	Variables Measured[a]	Health Criteria	
										Admission[b]	Analysis[c]
Minnesota	1947	1977	Professional and business men	281	M 45-54	Annual	1 day	?	Anthropometry Behavior and Personality	Yes	Yes
Duke I	1955	1976	Community Residents	260	M & F 60-94	2 yr	2 days	11	Psychiatric Psychology Physiology Anthropometry Blood chemistry Social history	Yes	Yes
Duke II	1968	1976	Community residents selected from register of health-insurance plan	502	M & F 45-69	2 yr	1 day	4	Psychology Social History Personality	No	Yes
Normative Aging	1963	Continuing	Community residents in Boston area	2032	M 25-75	5 yr	3 half days over 5-year period	3	Biochemistry Special senses Anthropometry Psychology Sociology	Yes	Yes
1000-Aviator	1940	1970	Cadets and officers in flight training	1056	M 20-30	Irregular	—	4	Physiology Psychomotor Psychology	Yes (rigorous)	Yes
NIMH	1955	1967	Community residents in Philadelphia area	47	M 65-91	5 yr	2 wks	3	Psychiatric interview Cerebral physiology Psychological tests Social history	Yes	Yes

Table II.13. Selected Longitudinal Studies of Adults—(*Cont'd.*)

Study	Began	Ended	Nature of Sample	N	Sex & Ages at Entry	Test Interval	Test Period	N Repeat Cycles	Variables Measured[a]	Health Criteria	
										Admission[b]	Analysis[c]
Basel	1955	1965	Community-residing CIBA employees and retirees	121	M 8-85 (Most 26-56)	1-2 yr	1 day	2 to 5	Anthropometry Physiology Sensory tests P.W.V.	No	No
Bonn	1965	1976-77	Community residents in W. Germany	220	M & F 60-75	2-4 yr	1 week	5	2- to 4-hr interviews Intelligence Personality	Yes	No
Framingham	1948	Continuing	Community residents	5209	M & F 30-59	2 yr	1 day	10	Blood chemistry end points—CV disease	No	Yes
Tecumseh	1959	1969	Total community	8641	M & F Birth to 70+ (45% under age 20)	3 yr	—	3	Anthropometry Physiology Blood chemistry Activity questionnaire	No	Yes
ABCC Adult Health	1958	1972	Non-exposed residents of Hiroshima	12,123	M & F birth to 70+	2 yr	½ day	7	Anthropometry Physiology Blood chemistry	No	No
BLSA	1958	Continuing	Community-residing males[d]	1142	M[d] 17-96	1-2 yr	2 ½ days	21 (as of 6/81)	See Chapter IV	No	Yes

[a] All studies in table recorded histories and performed physical examinations.

[b] Were health criteria used to select subjects for admission to the study?

[c] Were health criteria used to determine whether individuals' values were used in analysis?

[d] Women have been recruited to the study since 1978.

CHAPTER III

Design and Operation of the Baltimore Longitudinal Study of Aging

HISTORY

The program on aging of the National Institutes of Health (NIH) was originally established at the Baltimore City Hospitals (BCH) in 1940 (Stieglitz, 1940; Shock, 1947, 1980). Studies of age differences in the performance of selected physiological systems (cardiovascular, pulmonary, renal, muscular, nervous, endocrine, etc.) had used as subjects men aged 60+ who were residents of a domiciliary home for the aged (Infirmary) located on the grounds of the BCH. Since admission to the facility was based primarily on socioeconomic need rather than on health status, the group included residents without apparent disease. Along with staff members, these subjects, after clinical screening to exclude individuals in whom evidence of specific disease could be detected, constituted the population for cross-sectional studies on aging in humans conducted between 1941 and 1958.

At the same time, the desirability of collecting data on non-institutionalized subjects was clearly recognized, as was the fact that a proper study of human aging would require the characterization of age changes in a wide array of physiological, psychological, and social variables across the entire adult age range. Even as late as 1955, most of the literature reporting the effects of age on physiological and psychological variables in man was based on measurements of differences between young adults, usually college students, and older subjects who were frequently residents of hospitals or homes for the aged. Some of the conclusions about human aging derived from such studies were defective or erroneous, since they had compared presumably healthy and active young people with old people who were neither necessarily healthy nor active. Nor did this type of cross-sectional assessment permit an appreciation of the likely influence of birth-cohort effects on the differences observed between the two age groups.

In recognition of the need for a far more systematic approach to the understanding of age changes in man, a plan for a longitudinal study was developed requiring the recruitment of reasonably healthy and active community-dwelling people of all ages, who would be willing to undergo thorough and repeated testing over a major portion, if not the remainder, of their lives. The success of such an enterprise would obviously depend on access to a population of subjects who felt so challenged by their participation that they would voluntarily continue to return for repeated testing.

Our chance meeting with Dr. W.W. Peter provided the necessary access to a group of scientists and educators who had homes or summer cottages at Scientists' Cliffs, a section of the western shore of the Chesapeake Bay located about 60 miles southeast of Baltimore. Dr. Peter, a retired United States Public Health Service officer, was struck by the lack of data on the effects of age in normal people. Moreover, he expressed the strong conviction that people like himself, still living independently, should volunteer to serve as subjects for the study of aging. Thus was born the Baltimore Longitudinal Study of Aging (BLSA).

On his own initiative, Dr. Peter first recruited his friends to serve as volunteer subjects for the BLSA (Norris and Shock, 1960). His circle of friends, however, though large in number, would not provide sufficient subjects to conduct meaningful studies. Hence a plan was instituted whereby each participant actively recruited others among his friends, neighbors, relatives, and colleagues. This plan yielded so many subjects that a waiting list was quickly formed, from which subjects were called as facilities for testing them became available.

During the early stages of the study only two to four subjects could be tested each week, but as facilities expanded more subjects were tested, so that by 1968 about 600 men were actively participating. At this time the aging program was moved into a new building, located on the BCH campus, which was financed by Federal funds and designated as the Gerontology Research Center (GRC) (Shock, 1968).

Since its inception in 1940, the gerontological research program in Baltimore has operated within the administrative framework of NIH, principally as a unit of the National Heart Institute (1948-1963) and of the National Institute of Child Health and Human Development (1964-1975). After the National Institute on Aging (NIA) was established in 1974, all research activities housed at the GRC, including the BLSA, were transferred to NIA, and they have since constituted NIA's intramural (in-house) research program.

OBJECTIVES

From its initiation, the BLSA has had as its goal the systematic description of the processes of aging in humans. In pursuit of that broad goal the study has identified and sought certain objectives:

To describe, both cross-sectionally and longitudinally, the physiological and psychological effects of aging in persons who live in their communities.

To collect serial observations over a period long enough to permit statistically reliable calculation of rates of change in specific variables in individual subjects, and to identify individual patterns of age changes.

To include a broad spectrum of tests applied to the same subjects to determine relations among variables that might lead to an answer to the question, "Is there a general aging process common to a number of physiological and psychological processes, or is aging the end result of multiple independent processes?"

To separate the effects of aging from the effects of physical and mental disease on selected aspects of physiological and behavioral performance.

To assess the influence of age on progressive changes associated with diagnosed diseases.

To examine the effects of critical events, such as loss of job, retirement, death of spouse, or changes in life style or habits, on subsequent tests of physiological and psychological performance.

To develop predictors of age at death and to determine risk factors not only for specific diseases but also for other definable end points, such as loss of mobility, loss of independence, and institutionalization.

To attempt to develop indices of physiological age.

OPERATION OF THE BLSA

To provide a firm basis for interpretation of the results, this chapter outlines the design and operation of the BLSA as it has developed over the years in response to the many advances in the concepts and methods of gerontological research that have occurred since 1958. Many decisions have had to be made on the basis of hunches and best guesses rather than of solid fact, and it has often proven necessary to tailor the study to fit the available resources. In retrospect, while it is obvious that some decisions could have been better, others made out of necessity have turned out to the advantage of the study. A good example is the continuing addition of new subjects to the study population as the study progressed, which has made possible sequential analyses of period and birth-cohort effects (see Chapter I).

1. Administration

Both the scientific and the operational administration of the BLSA have been gratifyingly collegial throughout its years of existence. During the early years, decisions affecting the experimental questions to be addressed were made by the Director, Dr. N.W. Shock, in close conjunction with others of the senior scientific staff of the BLSA. These deliberations, usually undertaken in a regularly scheduled meeting, focused not only on the feasibility and scientific merit of the question to be explored, but also on the justification for addressing it to the BLSA population. Not infrequently, especially in the case of proposed cross-sectional studies, it was determined that utilization of BLSA participants was not essential, and that such investigations could be accomplished as readily by the use of other subjects specifically recruited for the purpose.

Such deliberations emphasize the need for constant overview of the uses to which the BLSA population is put. A concerted effort has been made to avoid the introduction of tests or procedures that are unlikely to lead to longitudinal exploration, or that will unnecessarily add to the already crowded test schedule to which BLSA participants are exposed.

Since 1981, the overall operation of the BLSA has been overseen by a Steering Committee, which includes the NIA Scientific Director as chairman along with five senior investigators whose collective research interests and experience embrace both biomedical and behavioral aspects. This committee serves as final arbiter of all new proposals for BLSA research studies. It is also charged with assessment of the readiness of BLSA data bases for analysis, to decide whether certain tests should be terminated and others introduced.

Operational aspects of the study, including scheduling of visits and tests, as well as communication with the BLSA participants, are the responsibility of a central staff. The subjects in the longitudinal study have always been regarded as participants, never as "guinea pigs." Investigators meet subjects at every opportunity to discuss the goals and operation of the study, and one or more staff members lunch with the subjects each day of their visits. In the early years of the study senior staff members joined the subjects for dinner, and in recent years close association has continued. In addition, each subject receives a regular newsletter that lists publications based on data from the longitudinal subjects and calls attention to important topics being studied, lists subjects who have died, and describes projected studies.

Tests identified by staff physicians as inappropriate or hazardous for certain subjects are excluded from their schedules. Similarly, the selection of subjects for special tests or the decision to exclude them is the responsibility of the scientific staff. Exclusions that are not based on medical considerations usually reflect conflict among test procedures, stemming either from competing time constraints or from incompatibility among tests; an example is the interference of insulin-infusion studies with tests of pituitary function. These decisions serve as the basis on which the test schedule for each subject is developed.

2. Participants

Development of the study panel. It is often difficult to enlist non-institutionalized volunteers in health-related research. It is even more difficult when participation entails spending two days and nights in a hospital setting, when candidates must furnish their own transportation, and when they are asked to return periodically for additional testing. Certain test procedures may prove burdensome, those designed to assess cognitive functioning can be ego-threatening, and still others call for the expenditure of considerable mental or physical effort. Despite these many and varied demands, the enthusiasm shown by the original members and their efforts at recruiting others

GERONTOLOGY
RESEARCH CENTER
NATIONAL INSTITUTES OF HEALTH
PUBLIC HEALTH SERVICE
DEPARTMENT OF HEALTH
EDUCATION AND WELFARE

Figure III.1. BLSA volunteers Mr. and Mrs. William E. Wyman of Dover, New Hampshire, during a visit to the GRC.

Table III.1. Recruitment Pattern by Age, Cycles A-K

Cycle	Inclusive Dates	Age (yr) 17–19	20–39	40–59	60–79	80–96	Total
A	February 1958–June 1961		51	119	86	4	260
B	July 1961–June 1963		38	81	37	2	158
C	July 1963–June 1965	1	23	68	39	3	134
D	July 1965–June 1967	1	21	61	32	2	117
E	July 1967–June 1969		29	48	36	6	119
F	July 1969–June 1971		57	30	41	3	131
G	July 1971–June 1973	2	29	1	10	3	45
H	July 1973–June 1975		52	2	1	5	60
I	July 1975–June 1977	1	44	14	4	1	64
J	July 1977–June 1979		10	4	7		21
K	July 1979–June 1981	1	15	6	8	3	33
Total		6	369	434	301	32	1142

contributed materially to the successful initiation and continuation of the BLSA. Figure 1 depicts two of the volunteers who have made the study possible.

Over the years, a few subjects joined the program at the invitation of staff members, and some volunteered after reading or hearing about the study. More than 90% of participants, however, were proposed for membership by friends or relatives who were already taking part in the study. In consequence, most new members have at least some familiarity with the nature of the testing program by the time of their first visit, and because many participants elect to return to the GRC together their visits tend to be pleasant social occasions.

During the initial visit, the nature and goals of the study are explained to each subject, and special emphasis is placed on the fact that, while the program is not designed to provide medical services, the results of the clinical and laboratory tests will be referred to each participant's personal physician. When the study started, a letter describing the tests to be performed was sent to the physician, who was asked to report directly to the GRC if, in his opinion, any of the tests proposed would present a hazard to his patient. Since none of the physicians raised objections to any of the tests, the letters were discontinued in 1962.

All subjects sign a statement of informed consent, which specifies the nature of all procedures and tests to be used, indicates potential hazards, and clearly states that each subject has the right to decline participation in any of the tests. New test procedures are reviewed and approved by the BCH Institutional Review Board.

In order to facilitate data analysis the repeated observations were grouped in 24-month periods, or cycles, designated alphabetically. Table 1, which shows the recruitment pattern by age groups for cycles A through K, reveals that recruitment of new subjects was most active between 1958 and June 30, 1971. Table 2 reflects the age composition of the panel at the end of each cycle.

Table 3 shows the total number of subjects who had been tested by June 30, 1981, distributed by age at time of first visit and number of visits. By that date 1142 subjects

Table III.2. Panel Age Composition, Cycles A-K

		Age (yr)					Totals	
Cycle	Inclusive Dates	17–19	20–39	40–59	60–79	80–96	Subjects	Visits[a]
A	February 1958–June 1961		51	119	86	4	260	369
B	July 1961–June 1963		69	175	119	7	370	445
C	July 1963–June 1965	1	73	214	143	12	443	536
D	July 1965–June 1967	1	64	272	178	20	535	646
E	July 1967–June 1969		62	287	201	22	573	793
F	July 1969–June 1971		106	292	239	31	668	913
G	July 1971–June 1973	2	99	262	226	34	623	784
H	July 1973–June 1975		135	227	223	43	628	776
I	July 1975–June 1977	1	156	217	218	39	631	750
J	July 1977–June 1979		121	205	228	41	595	691
K	July 1979–June 1981	1	118	196	222	48	585	695

[a] The number of visits within a cycle exceeds the number of subjects because BLSA participants over 60 years of age visited GRC more than once during the 2-yr cycles.

Table III.3. Cumulative Number of Male Subject Visits by Age at First Visit[a] (as of June 30, 1981)

	Age at First Visit (Yr)									
Number of Visits	17–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90–96	Total
1	6	143	226	242	192	153	148	27	5	1142
2	4	128	202	217	175	136	121	20	3	1006
3	4	101	176	201	155	123	104	16	3	883
4	3	77	145	179	145	113	90	11	3	766
5	1	49	123	173	135	100	76	7	3	667
6	1	31	109	161	122	96	60	6	1	587
7	1	21	88	154	107	88	55	6		520
8		15	74	133	91	79	52	5		449
9		12	64	113	79	68	44	4		384
10		9	48	87	62	58	33	4		301
11		9	32	68	51	52	26	2		240
12		3	17	45	39	39	19	1		163
13		1	8	24	28	29	15			105
14		1	2	12	20	26	9			70
15			2	6	13	19	5			45
16			1	2	10	14	3			30
17			1	2	6	9	2			20
18			1	1	2	5	2			11
19			1		1	3	1			6
20						2				2
21						1				1

[a] As a guide to interpretation of this table, it may be noted, for example, that 122 of the 192 subjects who began their participation in the BLSA when they were between 50 and 59 yr of age have made 6 visits or more since their entry into the study. Similarly, 2 of the 27 subjects who were between 80 and 89 at entry have made 11 visits, and one of these 2 had a 12th visit. The last column indicates, for example, that 240 subjects of all ages have made 11 visits.

had been seen one or more times, and nearly half of these, 520, had made seven or more visits.

While no systematic effort has been made to identify the motives that have prompted subjects to volunteer, it is clear from casual conversation that the opportunity to obtain a comprehensive physical examination at regular intervals has been an important factor. Nevertheless, since virtually all participants have personal physicians to whom medical reports are sent, and most are able to afford medical care, the physical examination cannot be more than partial compensation for the many demands of participation. It is obvious that a desire to contribute to aging research has been a large part of the motivation.

Although in the early years of the study subjects were not selected on the basis of age, after July 1970 preference was given to subjects in certain age decades in order to maintain at least 50 subjects in each five-year age group from 25 to 85 (Tab. 1). Most of the subjects live in the Baltimore-Washington metropolitan area, although some, especially those who have retired, return from such states as Michigan, Florida, New Hampshire, Maine, Arizona, and California.

When the study was initiated the resources available to the GRC for overnight housing of participants consisted of a single room, so that bathroom facilities had to be shared with ambulatory elderly male patients in a large ward of the BCH. It was thus necessary to limit the longitudinal study to men. With the opening of the GRC building and renovation of the ward space in the hospital in 1968, physical facilities became available for housing women. Because of limitations of staff and budget, however, recruitment and examination of women did not begin until 1978.

Socioeconomic and educational characteristics and health status of the participants. The socioeconomic characteristics of the sample generated by this method of recruitment are those of an upper-middle-class segment of the general population. An assessment of the social characteristics of subjects recruited up to June 30, 1981, is presented in Table 4. As the table reveals, 84% of the subjects were identified at initial visit with professional, technical, and managerial occupations, 71% had bachelors' or higher academic degrees, and 73% (82% of those reporting) rated their financial situations as comfortable or better.

In addition, the sample includes a high proportion of government employees: 56% were or had been in municipal, state, or federal employment, exclusive of public-school teaching and military service.

The sample at entry (Tab. 5) was also relatively homogeneous with respect to reported health, race, marital status, and religion. That 85% (93% of those reporting) rated their health as good or excellent is unremarkable, since one of the objectives of the study was the investigation of healthy aging. No subject was denied participation for reasons of health; nor were criteria of race or social status applied to restrict admission to the program.

It is conceivable that men of different ages or generations might have volunteered for different reasons, with the result that the sample would have varied in its social composition by age. This source of bias seems largely to have been avoided, however, in that the socioeconomic attributes listed in Table 4, and the distributions of health, race, and marital status by age in Table 5, either reveal a substantial correspondence among age groups or reflect changes in financial, health, and marital status that are usually associated with older age groups.

Although above-average educational level and socioeconomic status were not

Table III.4. Percentage Distribution of Male Subjects by Age and Socioeconomic Characteristics at First Visit (as of June 30, 1981)

	Age at First Visit (Yr)					Aggregate (17–96)
	17–19	20–39	40–59	60–79	80–96	
Number of Subjects	6	369	434	301	32	1142
Occupational Class						
Professional		56.9	66.6	67.1	71.9	63.4
Managerial		18.4	23.0	18.3	21.9	20.1
White collar		10.0	4.6	7.6	3.1	7.1
Blue collar		13.8	5.8	7.0	3.1	8.6
Student	100.0	0.8				0.8
	100.0	99.9	100.0	100.0	100.0	100.0
Educational Status						
Less than baccalaurate degree	100.0	36.3	23.7	25.2	31.2	28.8
Baccalaurate degree		32.2	27.6	24.9	18.8	28.0
Master's degree		20.6	22.4	14.6	21.9	19.6
Doctoral degree		10.8	26.3	35.2	28.1	23.6
	100.0	99.9	100.0	99.9	100.0	100.0
Present Economic Status						
Can't make ends meet		1.9	0.5	0.3		0.9
Enough to get along		22.0	13.8	10.6	9.4	15.4
Comfortable	50.0	55.0	67.7	68.4	53.1	63.3
Well-to-do	16.7	5.4	9.0	16.3	9.4	9.8
No information	33.3	15.7	9.0	4.3	28.1	10.6
	100.0	100.0	100.0	99.9	100.0	100.0

specifically applied as criteria for admission to the study, the method of recruitment resulted in a population possessing both attributes. The participation of highly educated subjects offered a number of advantages for the study of aging: improved accuracy of such information as family history, past history of illnesses, current medical status, and medication being taken; accuracy of activity history; accuracy of nutritional diaries and ability to follow detailed instructions in recording dietary intakes; ability to understand and execute a variety of questionnaires; and increased probability that the subject would be interested in the scientific questions addressed by the study. In short, this selection of subjects provided the opportunity to study the effects of aging under conditions in which economic status and educational level would not seriously limit medical care, nutrition, and other factors affecting health over the life span. The results of the study may thus be considered to reflect the effects of aging under optimal socioeconomic conditions.

Stability of the sample and losses from the program; drop-outs. In spite of additions to and losses from the study population over the history of the program, the social characteristics of the sample have remained surprisingly stable. Table 6 compares groups tested during each two-year cycle with respect to various attributes at entry into

Table III.5. Percentage Distribution of Male Subjects by Age and Personal Attributes at First Visit (as of June 30, 1981)

	Age at First Visit (Yr)					Aggregate (17–96)
	17–19	20–39	40–59	60–79	80–96	
Number of Subjects	6	369	434	301	32	1142
Health Status						
Poor		0.5	1.2	0.7	3.1	0.9
Fair		3.0	5.5	9.6	6.2	5.8
Good		33.3	47.2	50.5	40.6	43.2
Excellent	66.7	50.1	39.4	35.5	21.9	41.5
No information	33.3	13.0	6.7	3.6	28.1	8.7
	100.0	99.9	100.0	99.9	99.9	100.1
Marital Status						
Never married	100.0	17.1	2.3	1.7	3.1	7.4
Married		77.8	94.9	89.0	53.1	86.2
Formerly married		5.1	2.8	9.3	43.8	6.4
	100.0	100.0	100.0	100.0	100.0	100.0
Religious Affiliation						
Roman Catholic	33.3	22.8	14.7	8.0	3.1	15.3
Jewish		7.0	7.6	3.0	3.1	6.0
Protestant	50.0	59.1	72.6	81.7	84.4	70.8
Other	16.7	8.1	4.1	6.6	9.4	6.3
No information		3.0	0.9	0.7		1.5
	100.0	100.0	99.9	100.0	100.0	99.9
Color						
White	100.0	95.1	96.6	98.3	96.9	96.6
Black		4.9	3.2	1.7		3.2
Other			0.2		3.1	0.2
	100.0	100.0	100.0	100.0	100.0	100.0

the program. Minor changes have occurred during the 23 years. Mean age rose from 53.0 years in Cycle A (1958–1961) to 57.0 years in Cycle K (1979–1981). Over the same period the proportion of subjects identified with professional, technical, and managerial occupations fell from 91% to 84%, and Protestant affiliation from 79% to 71%. Sample composition remained remarkably uniform over all cycles in the proportions of participants with bachelors' or higher degrees at first visit.

Although most subjects who joined the study were aware of its demands, some found it difficult to continue. While no final assessment of the causes of their withdrawal has yet been made, predictors include the status of the recruiter (whether the person who recruited the subject is still in the program), personality traits reflecting critical dispositions—both specific complaints about tests or staff members and more

Table III.6. Percentage Distribution of Male Subjects by Mean Age and Selected Attributes at Each Cycle[a]
(Total N in Parentheses)

	Cycle										
Characteristic	A	B	C	D	E	F	G	H	I	J	K
Occupational Class											
Professional, technical, or managerial	91.2 (260)	90.3 (370)	89.2 (443)	88.0 (535)	86.7 (573)	85.0 (668)	84.8 (620)	84.6 (604)	84.1 (603)	83.9 (571)	84.1 (552)
Education											
Bachelor's or higher degree	79.6 (260)	78.9 (370)	79.7 (443)	78.3 (535)	77.5 (573)	77.7 (668)	77.7 (623)	75.3 (628)	74.0 (631)	74.0 (595)	72.7 (585)
Health (Self-rating)											
Good or Excellent	93.7 (253)	94.1 (358)	93.0 (428)	92.6 (516)	93.4 (545)	93.0 (632)	93.8 (592)	94.1 (580)	95.8 (575)	95.1 (551)	95.7 (532)
Religious Affiliation											
Protestant	78.9 (256)	78.2 (327)	77.5 (440)	76.6 (533)	78.5 (572)	74.5 (667)	75.3 (619)	73.4 (601)	72.1 (599)	71.0 (567)	70.7 (549)
Mean Age	53.0 (260)	53.8 (370)	54.6 (443)	55.7 (535)	56.3 (573)	55.8 (668)	56.6 (623)	55.6 (628)	54.9 (631)	56.7 (595)	57.0 (585)

[a] Attributes other than age determined at first visit only.

diffuse ones about the program and its environment—and changes in personal circumstances such as illness of the subject or of members of his family.

When subjects who had not returned after one or two visits by June 30, 1977, were compared with respect to age, occupation, highest degree, religious affiliation, and financial status at first visit with those who had failed to return after three or more visits, no significant differences between the two groups were observed. The drop-outs as a group (N = 280) were also found to be similar, except in the level of academic degrees, to subjects who had made three or more visits and were still active as of June 30, 1977 (N = 658). Fewer drop-outs than active subjects had held masters' degrees and doctorates at first visit.

As of June 30, 1977, 1088 subjects had been tested at least once. At that time a major follow-up study was made of the participants who had formally withdrawn or had not returned in a three-year period. Table 7 summarizes the results. By the cut-off date, 150 subjects had died and 280 had failed to return within three years. Of the 280 all but five were located. Of the 275 subjects located, 56 had died since their last visit. A priority cascade was established for the 219 who were still alive. Our top priority was to persuade them to return for additional visits. Since several of the major reasons for dropping out—serious illness, institutionalization, distant move—made return difficult or impossible, we were pleased that 76 (35%) returned. As the next step in our follow-up, we asked that a home visit be permitted, and were rewarded with a further 19%.

Table III.7. Follow-up Study of Male Subjects Seen at Least Once from February 1958 to June 1977

Total subjects seen	1088			
Active as of June 1977	658			
Died while an active participant	150			
Failed to return within 3 years of last visit ("Drop-out")	280 →	280		
Could not be located		5		
Located by follow-up		275 →	275	
Died after leaving study			56	
Alive at time of follow-up			219 →	219
Re-enrolled in study				76
Visited at home				42
Responded to mailed questionnaire				71
Responded to telephone interview				9
Subject undecided				19
Refused further cooperation				2

Eighty subjects (37%) completed a written questionnaire or responded to a telephone interview with respect to their health status. There is still the possibility of some follow-up on 19 subjects (9%) who have not yet decided how to respond. Only two have refused further cooperation. We thus know whether the subject is still living in 99.5% of the cases, and health and social information is available in 97.6%.

3. The Tests

This section relates the administration of tests to the purposes and design of the study. A detailed description of the tests administered is presented in Chapter IV.

Criteria for selection of tests. The battery of tests used in the study was designed to cover a broad spectrum of physiological, psychological, and social variables in order to permit the evaluation of relations among them and their association with aging. In addition to tests of the performance of specific organ systems, such as the heart, lungs, kidneys, and muscles, tests of the ability to integrate the activities of a number of organ systems in adapting to physiological stresses, such as exercise, were included. Other tests, such as glucose tolerance, measured the rates of displacement and recovery of physiological variables.

Selection of tests for repeated administration was based on the following considerations: a) the probable importance of the variable in understanding age changes in performance and adaptability; b) evidence from cross-sectional studies that age differences are present in the function; c) established validity and reliability of the test; d) the presence on the staff of an investigator interested in the test and qualified to administer it and interpret the results; and e) the presence of minimal risk and unpleasantness to the subject.

Testing schedule. At each visit subjects spend 2½ days at the GRC. Meals are provided in the cafeteria of the BCH, but subjects frequently elect to take one of the evening meals outside the hospital. The length of visit was chosen to provide sufficient time for a diversity of studies, including tests such as basal metabolism and glucose tolerance that must be made under basal conditions. The visit schedule also permits the collection of 24-hour urine samples required for the determination of creatinine clearance, the excretion of hormones, etc., as well as the introduction of lengthy and complicated laboratory procedures.

During each visit subjects rotate through various laboratories on a schedule that is prepared in advance for each subject. Individual testing schedules are also based on the time that has elapsed since the performance of other tests that might produce interference with the results. Tests administered during the course of the study fall into four general categories:

1) Tests administered to all subjects on all visits. Examples include medical history and physical examination, anthropometry, basal blood pressure and ECG, vital capacity, maximum breathing capacity, creatinine clearance, and blood chemistry (fasting cholesterol, triglycerides, and fasting glucose).

2) Tests administered regularly to all eligible subjects, but not at each visit. Some tests cannot be administered together in the same visit because of interference caused by the administration of drugs or hormones, the amount of blood withdrawn, or the time required. The frequency of administration of particular tests was determined on the grounds of availability of investigators and subject time, interference with other tests, and length of the period over which age change might be expected to occur.

3) Tests administered at only one visit, primarily to assess stable characteristics of the subjects, include determinations of such genetic markers as blood groups and taste sensitivity to phenylcarbamate.

4) Tests introduced for specific approved investigations that, although not necessarily longitudinal in design, could be carried out with maximum efficiency and minimal cost because a great deal of background information about the subjects from the longitudinal study was available. The study of age differences in the physiological and psychological effects of ethanol is an example.

During the early years of the study subjects traveled to Baltimore in pairs. At first only two pairs of subjects could be tested each week, but as resources expanded the number grew with them, so that by June 1968, 12 subjects were scheduled each week: six to arrive at the GRC each Monday and six each Wednesday morning. Subjects arriving Monday morning left at noon on Wednesday, and those arriving Wednesday morning left at noon on Friday. In 1978 the schedule was altered so that 15 subjects were scheduled each week: Six arrived for testing on Monday, three on Tuesday, and six on Wednesday mornings. Each group remained at the Center for two days. (This schedule was followed in order to provide time for testing women in the program that began in January 1978.)

In the original design, subjects aged 20 to 70 years were tested at 18-month intervals, while subjects over the age of 70 were seen annually. In July 1970 the test interval was set at 24 months for subjects aged 20 to 59, 18 months for subjects aged 60 to 70, and one year for all subjects over 70 years.[2] Table 2 shows both the number of subjects tested in each cycle and the total number of visits. When subjects were tested twice in a cycle, only the data from the first visit were used for cross-sectional analysis, although the data from both visits were used in longitudinal analysis.

---

[2] Since July 1981, all subjects have been seen each two years.

Quality control of test administration. Responsibility for training and supervision of technical assistants, as well as overall quality control of the data collection, rests with each investigator, who also maintains files of original data and reviews summary data before they are introduced into the central data file.

Changes in methods. Improved methods for tests already in use are introduced when they become available. Whenever methods are changed, measurements are made at the same visit by both the old and the new techniques in a series of subjects distributed over the total age span, so that if systematic differences are found in the values obtained by the two methods, appropriate adjustments can be made in the earlier measurements to maintain comparability throughout the longitudinal analysis.

Introduction and elimination of tests. New and expanding interests of investigators, as well as new developments in the field, have led to the incorporation of new tests as the study has progressed. New tests are introduced only after review by the Steering Committee of a written proposal submitted by the investigators. The testing cycle and the age span of the sample make it possible to assess the potential value of a new test by cross-sectional analysis of data collected over a two-year interval.

Over the years of the study, some tests have been eliminated because the study had been completed, because the principal investigator had left the GRC, or because it seemed unlikely that further testing would yield fruitful results.

APPROACHES TO INTERPRETATION OF DATA

1. The Aging-Disease Dilemma

There is common agreement that, in order to characterize true biological age changes in specific physiological systems, measurements in subjects with known diseases or pathology of the organ system under study should be excluded from analysis. Hence the necessity that normal aging be studied in "healthy" or "disease-free" subjects. This requirement has not so far been satisfied. In the first place, for certain conditions it is difficult to certify that a given characteristic is absent; the scientific method can only show that with current methodology a given condition cannot be detected. Since disease may be present but undetectable, it must be assumed that new and more sensitive methods could show its presence.

Unlike some scientists who believe that there is no unique biological process of aging distinct from pathological or disease-related processes, BLSA scientists have adopted the working hypothesis that aging is distinct from disease. Under this assumption great care is taken in the refinement of raw data and the exclusion of data obtained from subjects with specific diseases. An example of this process may be found in our study of creatinine clearance *(Rowe et al., 1976b);* the numerical impact is shown in Table 8. Consider the following example: A major goal of studies of blood pressure currently under way is to describe changes that occur with aging, but to eliminate changes induced by drugs or diseases that would, in themselves, change blood pressure. Variables known or believed to influence either cardiac output or peripheral resistance were considered. Thus, if a subject reported that he was taking any of the following medications, measures from that visit were excluded: diuretics and antihypertensives, amphetamines, methylphenidate (Ritalin^{R}), digitalis, long-acting coronary, cerebral, peripheral, or "general" vasodilators, anti-arrhythmics, oral and injectable adrenal or ovarian steroids, certain non-steroid anti-inflammatory agents, major

Table III.8. Number of Subjects Excluded from Normal Group by Disease Category and Age Group[a]

	Age (Yr)			
	17–24	25–44	45–64	65–96
Nephrolithiasis		8	47	21
Urinary tract infection		17	32	25
Gout		4	10	4
Prostatectomy		2	11	60
Congestive heart failure			2	3
Coronary heart disease		7	45	44
Cerebrovascular disease		2	16	18
Diabetes mellitus		10	36	18
Abnormal urinalysis	1	10	11	16
Miscellaneous renal disease	1	15	17	18

[a] Since some subjects had more than one disease, the total number of individual exclusions (531) in this table is necessarily larger than the number of subjects excluded (336).

From *Rowe et al. (1976b)*

tranquilizers, ergot alkaloids, narcotic analgesics, and L-dopa. Lesser drugs, e.g., minor tranquilizers, antihistamines, and sedatives, have not been shown to affect blood pressure.

The presence of any of the following conditions or diseases leads to exclusion from studies on the relation between age and blood pressure:

- Coronary artery disease. The point system used is described in Chapter IV. Measures are included in age analyses until the diagnosis of coronary artery disease is established, whereupon all subsequent measures are excluded.
- Aortic valvular stenosis or incompetence.
- Definite evidence of cerebral infarction.
- Hypo- or hyper-thyroidism at the time of the visit.
- Diabetes mellitus. Subjects who have ever taken either insulin or oral hypoglycemic agents are excluded. Also excluded are subjects with fasting plasma-glucose values of 140 or more on two or more occasions.
- Cancer with evidence of systemic effects, such as weight loss or anemia.
- Renal diseases, such as polycystic or horseshoe kidney, nephrectomy, or laboratory evidence of uremia (see below), regardless of etiology.
- Miscellaneous diseases, including Fabry's Disease, panhypopituitarism, and Buerger's Disease. Our study included no known cases of pheochromocytoma, adrenocortical disease, or coarctation of the aorta.

There are also laboratory criteria for the exclusion of blood-pressure measurements from estimates of the effects of aging:

- Abnormal urinalysis.
- Twenty-four-hour creatinine clearance with age-adjusted centile ranking of less than 1%, confirmed by age-adjusted serum-creatinine concentration ranking of less than 1%.

- Hemoglobin of less than 11 g/dl, or hematocrit of less than 33% or more than 55%.

One further exclusion is made, not on the basis of drugs, diseases, or laboratory findings, but of the well-known stress effect of the initial blood-pressure reading, the "first-visit artifact" (which in some studies has extended to the second visit as well). An analysis for this variable in our study showed the stress-effect to be evident only on the first visit. Measurements made on the first visit are excluded from calculations of the regressions of blood pressure on age.

2. The Concept of Clinical "Clean-Up"

The process of identifying and excluding subjects with conditions or diseases that might influence the values of a variable under study, or of excluding certain data points, is referred to as clinical clean-up. The necessity of a clinical clean-up preliminary to the analysis of data on aging processes is controversial. Consider the following example:

The creatinine clearance of a subject has been measured annually as he aged from 60 to 70 years. The slope of the 24-hour creatinine clearance on age, computed at age 70, is a measure of the rate of aging of his renal function, or glomerular filtration rate. After his last visit, certain symptoms develop; a nephrectomy is performed to remove a renal carcinoma. When he returns at age 71 a marked drop is of course found in the measured creatinine clearance. If the goal of the study is to define the effects of normal aging processes on selected physiological functions, should this last datum be included? Should a new slope be computed on data collected from age 60 to age 71, or should the assessment of the aging of renal function in the subject be stopped at age 70? In this example, it would appear absurd to disregard the surgical removal of a kidney—or to classify the event as normal aging. Conclusions based on computed slopes of decline that do not take this kind of event into account are bound to be misleading.

But consider a second example. Suppose that investigators studying renal function believe that hypertension may accelerate decline, and that subjects with this disease should be excluded. Since blood pressure increases with age, if a standard cut-off such as systolic pressure above 140 is adopted, the number of very old subjects eligible for inclusion in the study will be severely limited. Moreover, those who remain are likely to represent a "biological élite" whose health status is extraordinary. It is difficult to argue that normal aging can be studied in such a supernormal population. Utilizing a higher, "age-adjusted" cut-off point for older subjects is also unsatisfactory: If systolic pressures above 140 adversely affect the kidneys of 40-year-olds, surely they must adversely affect the kidneys of 80-year-olds. In this example, clinical clean-up may be inadvisable.

There are a number of alternative decisions the investigator may make, depending on his research goals. He may, for example, decide to screen for some medical conditions (such as surgical removal of a kidney), but not for others (such as hypertension), on the presumption that the former are both more directly relevant to the organ system under investigation and less likely to represent normal aging changes.

If the goal is the detection of pure aging effects, then a rigorous screening, leaving only a select group, may be warranted. Results from these studies are particularly useful in establishing that declines in functioning are an intrinsic part of the aging process.

Failure to find age changes would be difficult to interpret, however, since a true decline may be missed because of the bias introduced by the élite sample.

If an investigator is concerned primarily with the generalizability of his findings—if he wishes to describe changes that typically accompany age in the general population—then the best decision may be not to screen at all. This is also the strategy to be preferred if the researcher believes that the variable of interest does not change with age. If he can demonstrate stability despite both aging changes and the illnesses that accompany age, he has made a stronger case than could be made from a screened population.

At a minimum, it is our conviction that the decisions taken by each investigator must be presented in sufficient detail so that results will be properly interpreted and that comparability with other studies may be maintained. There is little question that other investigators faced with the same imperatives of categorization would arrive at a set of rules at some variance from ours. Indeed, candor requires us to admit that our own rules have tended to evolve not in response to an analysis of the end results of different strategies but according to standards of reasonableness (or unreasonableness) that have become evident as the initial set of rules has been applied. To illustrate: In an analysis of blood-pressure changes with age, a preliminary decision was made to consider the presence of a single cast of any type in the urine as indicative of a state of "renal disease" sufficient to exclude data from that subject's visit as "nonhealthy." This turned out to be nonsensical; urinary microscopic reports are replete with findings of single casts, with no ancillary evidence at all of renal disease at that visit or at prior or subsequent ones. When this criterion was modified to two or more casts, the percentage of excludable results dropped precipitously.

Perhaps the best strategy for dealing with this dilemma is to conduct and report analyses both with and without clinical clean-up. A comparison of the results of the two approaches will provide a better basis for both interpretation and generalization. If the same results are found, we can infer that the effect should be considered a manifestation of "aging," and also that it is found in the population in general. If results differ, we must be much more cautious in interpreting them, because either disease processes or sample biases could account for the effects. At this point it might be useful to conduct analyses on subgroups distinguished by particular diseases, to see if a single category of illness can explain the discrepancy. For example, screening for surgical removal of the kidney might make a considerable difference, whereas screening for hypertension might not. In these analyses we could learn not only about aging, but also about the consequences of specific disease states. This information, in turn, might be useful to future investigators who want to know which conditions are most important in a clinical clean-up.

THE DATA-MANAGEMENT SYSTEM

The data-management system of the BLSA provides a flexible facility for collection, storage, and analysis of those physiological, psychological, biochemical, medical, sociological, nutritional, and body-composition measurements described in Chapter IV. The system accommodates a variety of input modes and provides a wide-ranging, continually changing set of summaries and analyses responsive both to the needs of scientists and managers and to the legal requirement that confidentiality be assured. Scientists require access to a variety of statistical packages in addition to the implementation of their personal methodologies. Manager summaries, reports, projections, and schedules in most cases require programming *de novo.*

Facilities include a Digital Equipment Corporation (DEC) PDP 11/70 computer, which utilizes MUMPS (Massachusetts General Hospital Utility Multi-Programming System) as an operating and programming system and serves as a secure repository of the BLSA data. Scientists and managers gain access to this system through remote terminals in the laboratories and offices for entry, review, editing, and retrieval of single-record data.

The data base comprises 670,000 records, of which 380,000 are in fixed-field format (data in identifiable locations that remain the same for all persons and times). Fixed-field records include scientific results and numerically encoded data from medical histories and physical examinations. Medical diagnoses are recorded on another 60,000 records. The remaining 230,000 records contain verbal responses to specific questions, as well as free comments on fixed-field information. All free-form data are retrievable through numeric subject-matter codes that precede each entry in these fields, and by key-word search.

Batch-processing support for computation and analysis of the data includes a local Digital Equipment Corporation (DEC) VAX 11/780 computer and NIH's Division of Computer Research and Technology in Bethesda, Maryland, some forty miles distant, access to which is provided by a remote job-entry system (Data General Eclipse C-150) at the GRC.

Software is available to extract subsets of the data from the data-base system in forms suitable for analysis by standard packaged statistical programs (SAS, BMDP, SPSS, etc.) or by programs developed specifically for the BLSA by GRC scientists. Examples of the latter include several scanning routines that may be applied to analysis of data from subjects who have made several visits. Trends with time in a given variable, significant deviations from the trend or mean performance level, and regressions on time may all be evaluated by these routines and are essential both to quality control and to preliminary longitudinal analysis.

When data are available for only one or two visits, another program permits comparison of the subject's performance level with an appropriate distribution of similar values. Each subject must be compared with others measured at the same time, of similar age, of a similar disease classification, of similar body composition, height, or weight; or with subjects from another study in the literature in which the distribution of values is sufficiently well described.

Laboratory variables may be plotted against age, height, smoking habit, disease class, or any other variable in the file. This provides a quick and easy comparison of variance, for example, between old and young, tall and short, smokers and nonsmokers, etc. As a tool for cross-sectional analysis, these plots permit a judgment about the appropriateness of a linear or other fit to the data. Longitudinal data can be evaluated by comparison of these relations from one time of the study to another. A variation among the relations shown in the plots of a variable for different times suggests the need for a review of the laboratory results to determine whether longitudinal change, secular trends, or measurement error is responsible for variation over time.

Age regressions are based on a selected number of serial observations for a variable, such as the first five, all available, etc. Individual regressions are summarized in one or more age groups for the purpose of presenting the changes in subjects 35–44, 40–49, 40–59 years of age, etc. Plot programs are available for displaying age-group slopes around mean x, y, with a line length equal to the average period of observation on the time axis.

More specific and sophisticated computer routines for special data analyses, e.g., multivariate analysis, are developed by GRC investigators as the need arises.

CHAPTER IV

Tests Administered

INTRODUCTION

The Baltimore Longitudinal Study of Aging (BLSA) is unique among longitudinal studies in the large number of tests that have been administered serially in individuals ranging in age from 17 to 103 years. Observations include medical, genetic, biochemical, physiological, and behavioral variables. The comprehensive nature of the data base makes possible systematic description of the relations among aging trends in various types of human performance that may increase our understanding of the mechanisms of aging.

This chapter describes the tests administered and the procedures followed. Although some of the tests and the conditions under which they were administered have been published, many have not. Some of the tests and procedures are thus described in detail, while others are documented by references to pertinent articles in the literature. Information is provided about methods, the periods when the tests were administered, and the subjects who were tested. Table 1 lists the tests administered, as well as the number of data points for each as a function of number of subject visits. All data reflect our cumulative experience through June 30, 1981.

CLINICAL EVALUATION

1. History

In advance of each visit to the Gerontology Research Center (GRC), the subject is sent a medical-history questionnaire to complete and bring to the Center. During the interview, the examining physician reviews the information with the subject and if necessary supplements it. The history is further supplemented, when indicated, by summaries of hospital or physicians' records.

2. Physical Examination

At each visit a history and physical examination are recorded by staff members of the GRC on standard forms. The physicians, who usually serve two- or three-year National Institutes of Health fellowship appointments, are selected because of their interest in one of the research programs of the GRC. Most spend six to eight hours per month conducting physical examinations of the longitudinal subjects; the rest of their time is devoted to clinical and laboratory research in the operating section of the GRC to which they have been recruited.

When special diagnostic tests are necessary, they are conducted in the clinical facilities of the Baltimore City Hospitals (BCH). All reports of physical examinations are reviewed by a senior staff physician. After review, the clinical and laboratory data are entered in the central data bank and a summary report is sent to the subject's private physician.

Table IV.1. Tests Administered

		Number of Subjects Tested			
	Total No. of Tests	One or More Times	Two or More Times	Three or More Times	Four or More Times
Medical History, Physical Examination, and Laboratory Tests	7395	1142	1006	883	766
Medical history					
Physical examination					
Urinalysis					
Hemogram					
Hematocrit					
Hemoglobin					
Red-cell count					
White-cell count—differential					
Serological test for syphilis					
Chest x-ray					
Resting electrocardiogram	7339	1141	1004	883	762
Cornell Medical Index	2243	1114	649	361	89
Smoking history	7391	1142	1006	882	765
Genetic Factors					
Family history	1142	1142			
Taste test (PTC)	977	977			
Blood typing	766	766			
Dermatoglyphics	682	682			
Lateral dominance	605	461	144		
Biochemical Tests					
Serum albumin and globulin	163	161	2		
Plasma cholesterol and triglycerides	4655	944	812	715	606
Plasma creatinine (see below, "Renal Function")	2049	761	629	438	155
Plasma testosterone	83	83			
Plasma pituitary hormone (AVP)	66	66			
Body Composition					
Anthropometry	7395	1142	1006	883	766
Standing measurements					
Biacromial and waist diameter					
Chest, waist, and buttock depths					
Elbow and neck circumferences					
Height					
Weight					
Skinfold thickness					
Bilateral back, arm, abdomen, and chest					
Chin					
Recumbent measurements	4188	1083	824	668	549

Table IV.1. Tests Administered—(*Cont'd.*)

		Number of Subjects Tested			
	Total No. of Tests	One or More Times	Two or More Times	Three or More Times	Four or More Times
Body Composition—(*Con't.*)					
Anthropometry—(*Con't.*)					
Recumbent measurements	4188	1083	824	668	549
Circumference at iliac crest					
Circumference at greater trochanter					
Length of humerus—right and left					
Circumference of arm at humerus mid-point—right and left					
Length of ulna—right and left					
Circumference of forearm at ulna mid-point—right and left					
Length of femur—right and left					
Circumference of leg of femur mid-point—right and left					
Length of fibula—right and left					
Circumference of calf at maximum—right and left					
Bone density and diameters					
Bone density					
Hand x-rays	2664	1028	697	464	274
Photon scanning	1026	601	317	94	13
Bone diameters from x-rays	2664	1028	697	464	274
Midshaft overall diameter (hand)					
Midshaft medullary diameter (hand)					
Midshaft CCT (hand)					
Metacarpal length (hand)					
Elbow					
Knee					
Wrist					
Ankle					
Body fat thickness from x-rays	377	374	3		
Lean body mass—Behnke Index	4095	1068	797	664	543
Basal metabolism	5672	1092	911	779	636
Creatinine excretion—24-hour	6595	1133	986	816	705
Body water					
Total body water (antipyrine dilution)	1008	717	236	51	4
Extracellular water (thiocyanate dilution)	1078	761	261	52	4
Body density (helium displacement)	132	132			
Nutrition					
Nutrition history	3794	960	791	638	519
Seven-day diet diaries	2755	845	614	489	362

Table IV.1. Tests Administered—(*Cont'd.*)

		Number of Subjects Tested			
	Total No. of Tests	One or More Times	Two or More Times	Three or More Times	Four or More Times
Neuromuscular function and exercise					
Tapping test	6228	1072	934	810	690
Reaction time					
Touch	1833	782	544	310	148
Simple and choice auditory	1332	675	396	178	61
Reflex time	1317	658	407	177	60
Nerve conduction velocity, motor	2234	834	626	402	237
Exercise screening tests					
Strength test	3697	970	808	603	456
Hand dynamometry—right and left					
Bilateral adductors, abductors and rotors					
Low-level crank-turning ergometer	3295	924	747	543	408
Maximum work rate—brief	3294	924	747	543	408
Submaximal exercise	630	466	151	13	
Metabolic response—O_2 & CO_2					
Ventilation					
Blood pressure and pulse rate					
Blood chemistry	200	178	22		
Lactic acid					
Growth hormone					
Activity history questionnaire	5042	1077	865	727	619
Renal Function					
Creatinine clearance—24-hour	4435	1078	875	661	542
Urinalysis (see above, "Medical History, Physical Examination, and Laboratory Tests")					
Pulmonary Function					
Spirometry	6430	1109	945	813	711
Vital capacity—standing and recumbent					
Inspiratory and expiratory reserve volume					
Forced expiratory volume 0.5 and 1.0 sec					
Expiratory flow rates					
Maximum breathing capacity					
Pulmonary gas distribution indices (nitrogen wash-out)	1438	681	477	189	69
Total lung volume					
Residual volume					
Pulmonary and chest compliance	42	42			
Chest x-ray (see above, "Medical History, Physical Examination, and Laboratory Tests")					

Table IV.1. Tests Administered—(*Cont'd.*)

		Number of Subjects Tested			
	Total No. of Tests	One or More Times	Two or More Times	Three or More Times	Four or More Times
Cardiovascular Function					
Resting electrocardiogram	7339	1141	1004	883	762
Exercise stress electrocardiogram					
Master two-step test	3408	868	687	536	414
Graded treadmill test	2283	771	588	411	262
Systolic time intervals	326	326			
Echocardiography					
One-dimensional	270	258	12		
Two-dimensional	473	473			
Thallium scan of heart (exercise)	356	356			
Gated blood-pool scan	61	61			
Responses to cardiovascular drugs	89	89			
Volume plethysmography	140	140			
His-bundle recording (non-invasive)	80	80			
24-hour ambulatory ECG	150	150			
24-hour ambulatory blood pressure	50	50			
Plasma catecholamines (exercise)	30	30			
Carbohydrate Metabolism					
Oral glucose tolerance test					
1.75 g/kg body wt	1248	871	309	58	9
40.0 g/m^2 surface area	842	562	280		
Intravenous glucose tolerance test	1223	767	366	78	11
Cortisone-glucose tolerance test	774	581	160	30	2
Tolbutamide response test	1088	718	305	59	6
Glucose clamp	484	334	108	34	11
Exercise—I-V glucose tolerance	88	88			
Immune System					
Skin Fibroblast Culture	434	344	90		
Perception					
Tonography	3489	935	729	588	456
Visual screening test	5721	1038	901	784	664
Acuity					
Color vision					
Visual fields					
Audiometry—pure and pulse tone	4264	1012	845	683	557

Table IV.1. Tests Administered—*(Cont'd.)*

		Number of Subjects Tested			
	Total No. of Tests	One or More Times	Two or More Times	Three or More Times	Four or More Times
Cognitive Performance					
Learning					
Serial	1785	905	606	274	
Paired associate	1739	891	589	259	
Southern California tests of intellectual abilities	1378	824	554		
Word fluency					
Associational fluency					
Ideational fluency					
Consequences					
Intelligence tests					
Army Alpha	2939	1055	674	395	212
Wechsler Adult Intelligence Scale (Vocabulary only)	2045	1039	645	285	76
Benton Visual Retention Test	2043	1038	644	286	75
Logical problem solving					
I	529	304	225		
II	476	305	171		
Concept problem solving	1300	841	459		
Memory and response time	334	334			
Immediate free recall					
Delayed memory (recall and recognition)					
Digit memory					
Dichotic listening					
Simple and complex decision time and accuracy					
Personality, Behaviors, and Traits					
Guilford-Zimmerman Temperament Survey	2171	1075	681	327	83
Burgess-Caven-Havighurst attitude questionnaire	2122	1100	608	313	80
Questionnaire on attitudes toward old people	1243	811	360	67	5
Personal interview—family and sex history	777	777			
Imaginal Processes Inventory	613	613			
Thematic Apperception Test	66	66			
Stress-and-coping interview	116	116			
Schedule of life events	394	394			
Eysenck Personality Inventory	394	394			
Daily events check list	401	401			
Well-being assessment sheet	407	407			
Profile of mood states	394	394			
Parent-child relations questionnaire	337	337			
NEO Inventory	348	348			
Defense-mechanism inventory	140	140			
Social desirability scale	275	275			
Coping					
Self-Interview	76	76			
Questionnaire	149	149			
NEO Rating Inventory	146	146			

3. Laboratory Tests

Oral/dental examination. Since 1978, the oral health status of approximately one third of the BLSA subjects has been evaluated (Baum, 1981a). Data collected include the number of decayed (coronal, cervical), missing, and filled teeth, and indices of gingival and periodontal disease (Ramfjord, 1967). Oral mucosal tissues are examined. Two dental bite-wing and one panographic x-rays are taken. A detailed history of oral-hygiene habits, dental treatment, subjective complaints, and nutrition is obtained. Oral motor function is evaluated by physical examination (Bosma, 1976; Baum and Bodner, 1983). Assessments of swallowing, tongue and labial postural function, and masticatory ability are included.

Gustatory function is estimated in two ways: a) by determination of detection thresholds for four taste qualities, sweet (sucrose), salty (sodium chloride), sour (citric acid), and bitter (quinine sulfate) (Weiffenbach et al., 1982); and b) by determination of suprathreshold taste intensity through a magnitude-estimation procedure for each testant (Bartoshuk, 1978; Cowart and Baum, 1981). Salivary-gland function is assessed by collecting unstimulated whole saliva (volume determination after expectoration) and 2% citric acid stimulated parotid saliva (Baum, 1981b). From parotid saliva collections the following information is routinely obtained: flow rate, protein by A_{215}, K^+, Na^+, Ca^{++} by atomic absorption spectrometry; inorganic phosphate (Chen et al., 1956); anionic proline-rich proteins and amylase (by immunochemical assays).

Urinalysis. Routine clinical urinalysis, including microscopic examination of urinary sediments, is performed at each visit.

Hemogram. Hemoglobin, hematocrit, white-cell, and differential counts are made on each visit.

Serological test for syphilis is made on the first and every subsequent fourth visit.

Total serum protein, albumin, and globulin. These analyses were made at each visit from 1958 to 1969 in venous-blood samples drawn after 12 hours' fasting (7:00 P.M. to 7:00 A.M.).

Cholesterol. All determinations are made on fasting blood samples. Serum concentrations of cholesterol have been measured at each visit since April 1962 by the method of Abell et al. (1952) as modified by Anderson and Keys (1956). The method of Blankenhorn et al. (1961) has been used in the initial serum preparation phase so that triglycerides and cholesterol can be assayed on the same serum specimen. Since April 1969 all sera have been assayed by BioScience Laboratories, Van Nuys, California; in addition, once each month one eighth of the monthly samples are assayed in the GRC laboratory so that the BLSA can maintain its own reference method as a continuing quality control. From these analyses conversion factors have been developed for the BioScience analyses *(Hershcopf et al., 1982).* The BioScience Laboratories utilized the analytical method of Kessler (1967) until July 1970. Since that time the method of Wybenga et al. (1970) has been used.

Triglycerides. Serum concentrations of triglycerides have been measured at each visit on fasting blood samples by the method of Blankenhorn et al. (1961). Analyses by BioScience Laboratories use the method of Kessler (1967). Monitoring by BLSA is the same as for cholesterol (see above).

4. The Cornell Medical Index Health Questionnaire

The Cornell Medical Index Health Questionnaire (Brodman et al., 1949, 1960) is administered on visits 1, 5, 9, etc. The subject responds to 195 questions about his symptoms, health habits, and family history.

GENETIC CHARACTERISTICS

1. Family History

A detailed family history is obtained on the first visit, entered in the central data file, and updated on each subsequent visit.

2. Genetic Markers

Since the following tests concern characteristics that are assumed not to change with age, they are administered only once, usually but not always at the first visit.

Taste Test. The ability to taste phenylthiocarbamate (PTC) is recorded during the first visit. PTC is administered on filter-paper strips placed on the subject's tongue.

Blood Types. Typing for ABO, MN, Rh, Kells, Kidd, and Duffy factors is carried out at first visit only.

Dermatoglyphics of the Hand. Finger and palm prints of all subjects are collected by the Faurot inkless method with the digits comfortably extended during printing (Plato, 1978). The prints are classified according to the methods of Cummins and Midlo (1943).

Handedness. Hand, foot, and eye preference are determined to infer hemispheric dominance. Single-handed, two-handed, foot, and eye function are determined by a series of standard tasks.

BODY STRUCTURE AND COMPOSITION

1. Anthropometry

As many as 37 anthropometric measurements are made on all subjects in the standing and recumbent positions at each visit. Standard techniques use tapes for circumferences and lengths, calipers for diameters. In addition to height and weight measured without clothes, eight diameters, 12 circumferences, and eight lengths are measured. Skinfold thickness is measured with a constant-pressure Harpenden caliper at nine standard sites (Edwards et al., 1955; Tanner, 1955).

2. Bone-Mineral Measurements

Bone density and diameters from hand x-rays. Radiographs of each hand are taken postero-anteriorly at an average exposure of 1.0 sec at 100 mA and 60 kVp without intensifying screens. All measurements are made on the second metacarpal bones as described by Garn (1954, 1961, 1970). Specifically, the total width, medullary width at the midshaft of the bone, and length along the longitudinal axis are measured.

Bone density from photon scan. The Cameron technique (Cameron and Sorenson, 1963, 1968) of bone-mineral analysis is used to determine the mineralization of the radius and the ulna in both arms. This technique passes a collimated beam of monoenergetic photons through the combination of soft tissue and bone in a limb, and the resulting attenuation is monitored with a suitable photon detector. The source and detector are moved across the limb; the resulting absorption curve can be related directly to the scan. The Norland-Cameron Analyzer, an automated instrument used since 1972 for the determination *in vivo* of bone-mineral content, provides direct digital readouts of both bone-mineral density and bone width without external calculation or manipulation of data.

3. Subcutaneous Fat

In addition to measurements of skinfold thickness, estimates of subcutaneous fat were made from soft-tissue radiographs. From 1958 to 1973, measurements were made at each visit. Each 7-x-17-inch film contains views of seven sites on the trunk and limbs. Radiographic procedures and techniques were those described and illustrated by Garn (1954, 1961).

Measurements of skin and fat combined were made by means of a Helios dial-reading caliper calibrated to 0.05 mm *(Borkan and Norris, 1977)*. Sites of fat measurements on the trunk were bony landmarks, such as the top of the greater trochanter. For measurements on the calf and forearm, the widest part of the limb was chosen. Fat measurements were made at the following locations on each film: anterior calf, posterior calf, medial calf, lateral calf, lateral to greater trochanter, lateral to top of greater trochanter, lateral to anterior-superior spine of iliac crest, lateral to top of iliac crest, lower part of thorax (lowest rib), medial arm, and lateral arm. The average value for reliability coefficients determined by duplicate measurements on 20 films was 0.95 *(Borkan and Norris, 1977)*.

4. Estimates of Components of Body Mass

Behnke index. A quantitative classification of body build devised by Behnke is based on 11 circumferences and eight diameters of the body. Factors are used to indicate numerically the degree of fatness, muscularity, and skeletal size (Behnke, 1961, 1963). Measurements were made every fifth visit from 1961 until 1972; since then, measurements have been made at each visit.

Muscle mass. The total amount of creatinine excreted in 24 hours, measured at each visit, is used as an index of muscle mass (see below, "Renal Function").

Body-water compartments—estimates of muscle and fat. The volume of distribution of antipyrine is used to estimate total body water (Soberman et al., 1949). Thiocyanate space is used as the estimate of the volume of extracellular water (Levitt and Gaudino, 1950). Subjects who report a history of significant allergic reactions are not tested. Subjects tested between 1958 and 1967 received an intravenous infusion of antipyrine (1.0 g) and sodium thiocyanate (1.396 g) in 50 ml of normal saline. Plasma levels of antipyrine and thiocyanate were measured in peripheral venous blood at 2, 4, 6, and 8 hours after the infusion began. Antipyrine was estimated by the method of Brodie and associates (1949), with modifications (Shock et al., 1963). Thiocyanate levels are estimated by the method of Bowler (1944). Volumes were estimated from the slopes of the linear regressions of the logarithm of the concentrations of antipyrine and thiocyanate in the four blood samples. Intracellular water, calculated as the difference between antipyrine (total body-water volume) and thiocyanate (extracellular water volume), was used as an index of fat-free tissue.

Body density—body fat. From 1959 to 1966, estimates of body density were made by the Siri technique (Siri, 1956), by which the subject sits in an airtight chamber into which a known amount of helium is introduced. The temperature and helium concentration in the air surrounding the subject were monitored continuously for 15 minutes, by which time equilibrium is achieved. The body volume of the subject was calculated from the helium concentration in the air surrounding the subject. Body density was obtained from the ratio of total body weight to body volume. With appropriate corrections for bone density, an estimate of body fat was made for each subject. Quality control was maintained by the introduction of standard carboys of known volume into the chamber for measurement each day (Norris et al., 1963; Siri, 1962).

Because of excessive operating costs and technical difficulties, this measurement was discontinued in 1966.

5. Strength Tests (see below, "Neuromuscular Function and Exercise")

6. Basal Metabolism

Basal oxygen uptake is determined at each visit by the open-circuit method described by Shock and Yiengst (1955). Three ten-minute samples of expired air are collected on each of two consecutive mornings after the subject's overnight stay at the GRC. Until 1965, aliquots of expired air were analyzed for O_2 and CO_2 content by the Haldane apparatus. Between 1965 and 1968, expired air was also analyzed for O_2 by the paramagnetic method (Beckman Paramagnetic O_2 Analyzer, Model G–2), and for CO_2 by infrared absorption (Beckman Model LB–1). Once the two analytical systems were shown to be equivalent, all subsequent analyses were done by the more modern methods. Both instruments are calibrated daily with standardized mixtures of O_2 and CO_2 obtained commercially in standard pressure tanks and checked by the Haldane method (Tzankoff and Norris, 1977).

This test also provides data on basal respiratory volume, CO_2 elimination, heart rate, and blood pressure. Blood pressure and heart rate are recorded by nursing personnel.

NUTRITION

1. Dietary Habits

From 1961 to 1965, and from 1968 to 1975, assessment was made of dietary habits of the subjects. On the first day of a visit, after a brief introductory review of his diet, each subject was instructed by a trained nutritionist in the keeping of an accurate dietary diary. Various plastic food models were used to teach subjects appropriate assessment of portion size. For practice, a trial diary kept during the subject's visit at the GRC was monitored by the nutritionist. A seven-day food record, begun on the subject's return home, was mailed to the nutritionist. From 1961 to 1965, a subset of the subjects collected seven-day records at three-month intervals. All records were verified by the nutritionist with the subject on his next visit to the GRC.

Data from each record were coded by the nutritionist and entered on tabulation cards for computer analysis of mean daily nutrient intakes, variances, and the major food-group sources of the nutrients by a computer program developed by the Heart Disease Control Program, Bureau of State Services, United States Public Health Service. The analysis identified average daily intakes of total calories, protein, total fat, saturated fatty acids, oleic acid, linoleic acid, linolenic acid, and other fatty acids, total carbohydrates, simple and complex carbohydrates, alcohol, cholesterol, fiber, calcium, iron, and selected vitamins (vitamin A, thiamine, riboflavin, niacin, pyridoxine, and ascorbic acid). In addition, certain derived variables were computed (percentage of calories from protein, carbohydrate, and fat, and the polyunsaturated/saturated fatty-acid ratio) (McGandy et al., 1966).

2. Nutrition History

A nutrition history was recorded in a personal interview by a trained nutritionist during the periods 1961–1965 and 1968–1975, in order to detect unusual dietary features that might not have been revealed by the dietary diary.

3. Pyridoxine (Vitamin-B_6) Status

These studies were carried out from 1971 to 1973 (Rose et al., 1976). Venous-blood samples were obtained one to two hours after breakfast. Vitamin-B_6 supplementation by intake of vitamin pills was established from the patient's history. Plasma pyridoxal phosphate (PLP) was determined by a modification of the method of Chabner and Livingston (1970). Erythrocyte glutamic-oxaloacetic transaminase (EGOT) activity was determined by AutoAnalyzer Method No. 3 (Technicon Instruments Corp., Tarrytown, New York). EGOT was also measured after *in-vitro* stimulation with pyridoxal phosphate.

NEUROMOTOR FUNCTION AND EXERCISE

1. Tapping Test

This test, designed (Welford et al., 1969) to evaluate speed and accuracy in a visually controlled motor task, records the time required to place pencil dots alternately on two targets drawn on paper. A number of separate tests of increasing difficulty are carried out at each session. In these trials subjects are instructed to complete the task as rapidly and accurately as possible. Three sets of targets are used. Each set consists of pairs of vertical lines requiring movements of 50, 142, and 402 mm. The target widths (distances between the two parallel lines) are 32, 11, and 4 mm. The nine combinations of movement distance and target width are presented in different orders to different subjects in such a way that the serial positions both of the conditions and of the transitions from any condition to any other are appropriately balanced. Subjects have one trial with each of the conditions for practice, followed by a second trial with each pair of targets; their scores are derived from the second. The test was administered on each visit from 1960 to 1981.

2. Oral Motor Evaluation (see above, "Oral/dental examination")

3. Reaction Time

Reaction time to touch. Foot-reaction times are measured after stimulation by touching the sole of the foot. The subject is instructed to plantar-flex his foot as soon as possible after the touch. Tests are performed with the subject in the dorsal recumbent position. The muscle-action potentials are led from small solder electrodes placed over the small muscles of the left foot through paired leads and a preamplifier to a dual-beam cathode-ray oscilloscope. The subjects are connected with ground through a saline-moistened pad placed around the leg just above the ankle. The action potentials are recorded photographically from the oscilloscope screen.

Tactile stimuli to elicit voluntary muscle reactions are provided by a plastic rod tapered to a dull point at one end. Contacts are placed in the handle of the rod so that at the instant of contact with the skin the beam of the cathode-ray oscilloscope is triggered. The reaction times are measured by the interval between the beginning of the cathode-ray beam and the beginning of the muscle-action potential. A time-calibration curve (10 msec) is recorded on the second synchronized beam from the output of a standard oscillator. For each subject, mean values for reaction time are calculated after exclusion of the extreme high and low values, defined as those separated from the remainder of the distribution by an interval of more than 15 msec (Hügin et al., 1960).

This test was administered on alternate visits from 1958 to 1974. Since 1974, it has been administered on each visit.

Auditory reaction time. Between 1961 and 1976, simple and choice reaction times were measured concurrently with electroencephalographic (EEG) recordings from the occipital region of the head (Surwillo, 1963a). The stimuli were tones (250 Hz and 1000 Hz) presented over a loudspeaker at a level set by each subject to ensure that they were clearly audible. The duration of each stimulus was 0.3 second. No forewarning was given of the stimuli, which were triggered by the experimenter when alpha waves were evident on the EEG. In simple reaction time, the task was to press a button on a hand-held switch as quickly as possible whenever either tone was presented. In choice reaction time, the task was to press the button only when the higher tone was presented.

In 1974, a study was initiated in which only simple reaction time is measured; both tones are used. EEGs are no longer recorded, and the stimuli are triggered by the experimenter at random intervals of 10 to 25 seconds.

Reflex time. Reflex times are measured for stimuli applied to the sole of the foot. Plantar flexor reflexes are elicited by a brief longitudinal scratch applied with a plastic rod to the sole of the foot in the area in which the shortest, most consistent responses are obtained. The length of this scratch is approximately 5 cm, the duration approximately 200 msec. The intensity of scratch necessary to elicit flexion potentials ranges from a mere touch to a firm stroke. In no case has the stimulus been judged painful by the subjects.

A dual-beam oscilloscope is used to record the application of the stimulus and the muscle-action potentials (see above, "Reaction time to touch").

Fifty or more stimuli of each type are administered to each subject at each sitting. Average values are calculated if nine single values are obtained from the records (Hügin et al., 1960; Magladery et al., 1958).

Plantar reflex times were recorded in all subjects on alternate visits from 1958 until 1974, and have been recorded on each visit since 1974.

4. Nerve-Conduction Velocity

Nerve-conduction velocity is estimated from the time elapsing between the percutaneous application of an electrical stimulus to the ulnar nerve and the motor response in the muscles of the hypothenar eminence and the linear distance traversed. Electrical stimuli (supramaximal square wave shocks of 0.1 msec duration) are applied to the left arm over three separate points of the ulnar nerve, at the wrist, at the elbow, and 5 cm below the axilla. The cathode of the stimulating electrode is always placed over the part being stimulated. The response, recorded on a cathode-ray oscilloscope, is the action potential evoked in the muscles of the hypothenar eminence by the electrical stimulus to the nerve. One recording electrode is placed over the bellies of these muscles and the other over the tendons. Latency of response (the time between stimulation and the onset of the action potential of the muscles) is measured to the nearest 0.1 msec. Distances along the skin surface between centers of the stimulating electrodes and the recording electrode are measured to the nearest 0.1 cm (Norris et al., 1953; Wagman and Lesse, 1952).

This test was administered on alternate visits between 1958 and 1974.

5. Physical Activity: Muscular Strength and Work

Physical-activity history. The average level of physical activity was initially estimated from a personal interview or questionnaire (McGandy et al., 1966), and latterly by the questionnaire, which covers specific activities at home, at work, and in recreation, as well as variations in activity patterns such as trips and seasonal sports. The amount of time spent in each activity is expressed as a daily average. Time assignable to seasonal activities and activities that are pursued infrequently is expressed as an annual total and divided by 365 to obtain the daily average. The total daily energy expenditures are calculated for each subject by use of the caloric values for each activity determined by McGandy et al. (1966).

Strength tests. Maximum grip strength of right and left hands is estimated with the Smedley hand dynamometer (Fig. 1). All subjects are tested at each visit.

A special instrument was designed and constructed to measure strength of adductors, abductors, dorsal rotors, and ventral rotors of the arms. With this apparatus, the subject is seated with a one-inch bar in front of him. The subject pushes, pulls, lifts, or depresses the bar, which does not move. The force exerted on the bar is recorded by strain gauges. The maximum force developed in the highest of three trials at each maneuver is used as the strength index. The test has been administered to all subjects on each visit since 1960.

Coordinated exercise—cranking. The subject lies on his back on a bed and turns an electrodynamic-brake ergometer crank (Fig. 2) with his arms to accomplish 500 kg m of work at the slow and easy rate of 135 kg m/min (Kelso and Hellebrandt, 1934; Tuttle and Wendler, 1945). The output of the ergometer, which measures the work rate continuously, is monitored for variability of performance of the assigned rate (Norris and Shock, 1957).

Maximum work output is also determined. In this test the subject turns the ergometer crank at his greatest effort (maximum rate) for ten seconds. Four different resistances are used. Maximum work rate achieved at each work load is recorded. The test has been administered to selected subjects at irregular intervals.

6. Physical Activity: Oxygen Uptake

Submaximal workloads. Submaximal work (500 kg m at 135 kg m/min) is performed with the arms on the electrodynamic-brake ergometer (Fig. 2). The subject remains in a recumbent position throughout the periods of rest, exercise, and recovery.

Expired air is collected through a Siebe-Gorman face mask in serial ten-liter

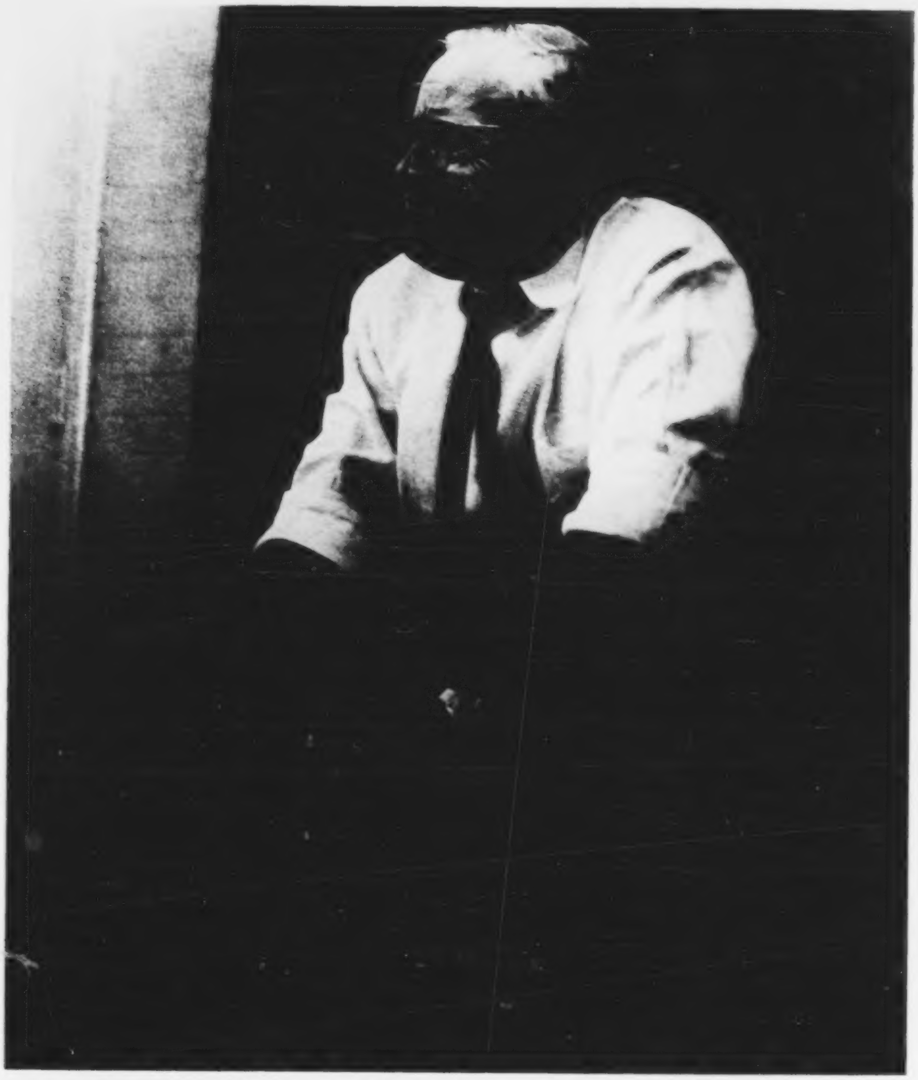


Figure IV.1. Measurement of hand-grip strength by Smedley dynamometry.

samples for 20 minutes before the exercise, during exercise, and for 40 minutes after exercise. The concentrations of O_2 and CO_2 in each ten-liter sample of expired air are determined by automatic gas analyzers (see above, "Body Structure and Composition," "Basal Metabolism"). All gas volumes are reduced to standard conditions (0°C, dry, and 760 mm Hg) (Carpenter, 1939). Pulmonary ventilation volume is determined by dividing the volume of air expired during the collection period by the duration of the period.

Rates of O_2 uptake and CO_2 elimination are estimated from ventilation rate and gas concentrations of the expired air by standard metabolism techniques (Peters and

Figure IV.2. Measurement of maximum work by bicycle ergometry.

Van Slyke, 1932). Maximum O_2 uptake, CO_2 elimination, and ventilation volume (l/min) are recorded from the highest value found for the volume in a single period during or after exercise. Total excess values for ventilation, CO_2 elimination, and O_2 absorption are obtained by summing over all periods during and following exercise the amount by which each individual metabolism period exceeds the value of the asymptote of the appropriate recovery curve. Net mechanical efficiency is calculated as the caloric equivalent of the work done, divided by the caloric equivalent of the metabolic cost of the work (total O_2 uptake, during and after exercise). The respiratory quotient is assumed to be 1.0. Oxygen debts are calculated as the amount of excess O_2 absorbed after the end of the exercise (Norris and Shock, 1957).

Measurements of blood pressure and heart rate are also made during the resting period and at 30-second intervals following exercise.

These tests have been applied to selected subjects at irregular intervals.

Maximum oxygen uptake. From 1975 to 1981, O_2 uptake was measured during graded maximal treadmill exercise in participants without clinical or stress electrocardiographic evidence of heart disease. According to the modified Balke exercise protocol (Balke and Ware, 1959), the individual walks at a constant speed throughout the test; the treadmill is elevated 3% every two minutes until the subject is unable to continue because of general fatigue, chest pain, shortness of breath, or leg discomfort. During exercise, the subject breathes through a mouthpiece connected to two large collection tanks by a flexible plastic hose. Expired air is collected for the second minute of each two-minute stage and analyzed on-line for O_2 and CO_2 content by a Beckman Medical Gas Analyzer LB–2. Correction of prevailing room temperature and barometric pressure permits calculation of the O_2 consumption in ml O_2 per min per kg body weight for each stage of exercise. The value at maximal exercise is generally considered the best available measure of physical fitness.

RENAL FUNCTION

1. Creatinine Clearance (24-Hour)

A non-fasting serum sample for the determination of creatinine levels is obtained from each subject on his arrival at the GRC, between 9:00 and 11:00 A.M. After the subject empties his bladder, a 24-hour urine collection is begun. A fasting blood sample for the determination of serum creatinine is obtained at 8:00 A.M. the next day; the mean of the two determinations is used in the calculation of creatinine clearance. Creatinine in serum and urine is measured as true creatinine, by a modification of the technique of Hare (1950). Clearance values are expressed as ml per 1.73 m^2 surface area (Rowe et al., 1976a).

In order to establish the validity of creatinine clearance as a measure of glomerular filtration rate, simultaneous inulin and creatinine clearances were performed in healthy male volunteers who were not participants in the BLSA. Creatinine and inulin determinations were made on the same blood samples drawn during a 45-minute intravenous infusion of insulin (Davies and Shock, 1950). The mean ratio of creatinine to inulin clearance was 1.29. Age had no significant effect on the ratio.

Creatinine clearance was determined on all visits from 1958 to 1974. Since July 1974 the test has been administered on visits 1, 2, and 5, and at every fourth visit thereafter.

2. Urinalysis (see above, "Clinical Evaluation")

3. Concentrating Ability of the Kidney

Water-deprivation tests were conducted from 1958 to 1962 *(Rowe et al., 1976b)*. There was no oral intake from 6:00 P.M. to 6:00 A.M.; after the subject emptied his bladder at 6:00 P.M., urine samples were obtained at 9:00 P.M., 12:00 midnight, and 6:00 A.M. for measurement of urine flow and osmolality. At 6:00 A.M. a plasma sample for osmolality and creatinine concentration was obtained. From these data solute excretion and osmolar clearances were calculated. This test was not repeated.

PULMONARY FUNCTION

All pulmonary-function tests were performed at various times of the day.

1. Spirometry

Earlier studies (1958–1962) of vital capacity and pulmonary subdivisions were performed with subjects both standing and recumbent; they inspired maximally and then expired maximally into a 120-liter recording Tissot spirometer. Complete spirograms were also recorded for recumbent subjects with a ten-liter closed-circuit spirometer. A mouthpiece and nose-clip were used for these collections. Three trials were made for each effort (inspiratory reserve volume, expiratory reserve volume, and vital capacity). The largest value was recorded as the value for the day. The resting tidal volume was determined by dividing the total volume expired in a ten-minute air-collection period by the number of breaths during this period as counted from a kymograph tracing (Norris et al., 1956). Since 1963, spirometric studies have followed methods and calculations described by Kory et al. (1961), including forced expiratory volumes ($FEV_{0.5}$ through $FEV_{6.0}$). These tests are performed at each visit.

2. Maximum Breathing Capacity

The maximum breathing capacity is determined in standing subjects who are asked to breathe as much air as possible into a spirometer through a low-resistance circuit for 15 seconds. Neither the rate nor the depth of breathing is specified, but the subject is urged to do his best throughout the test. The highest volume attained in three trials is taken as the value for the day (Norris et al., 1956). This test is administered on each visit.

3. Pulmonary Gas Distribution

Total lung volume and functional residual volume were determined by the nitrogen wash-out method using an open-circuit technique (Edelman et al., 1968). The studies were performed at various times of the day with subjects in the seated position. The gas supply and collection bags were enclosed in an airtight box. Tidal volume was monitored by a model 350 Servo-spirometer (Med-Science Electronics, St. Louis) connected to the box. Nitrogen concentration of gas sampled at the mouthpiece was measured with a model 300 AR Nitralyzer (Med-Science Electronics). The instrument was calibrated with five standard gas mixtures within the 2%–7% nitrogen range before each wash-out test. Continuous recordings of N_2 concentration and tidal volume were made with a model 1508 Visicorder (Minneapolis Honeywell Regulator Co., Denver). Subjects were allowed to accommodate to the apparatus while breathing air. A vital-capacity maneuver consisting of a full inspiration followed by a full expiration was performed during this period. After the subjects had returned to a steady ventilatory pattern (usually within 0.5–1.0 min), and at the end of a normal expiration, a seven-minute period of oxygen breathing was begun. Functional residual capacity (FRC) was calculated from the collected expired air. Corrections were made for inspired nitrogen concentration and tissue nitrogen excretion (Darling et al., 1940). Tidal volume was taken as the mean for the seven-minute period; anatomic dead space was estimated from the height of each subject (Hart et al., 1963).

Uniformity of ventilation was initially assessed by the use of the lung-clearance index (LCI) (Becklake, 1952), which was later supplanted by a new index less dependent on tidal volume (Edelman et al., 1968).

This test was administered between 1962 and 1979 to randomly selected subjects.

4. Lung and Chest-Wall Compliance

Since this test required the placement of an intra-esophageal balloon catheter to measure pressures, it was administered to only 42 subjects, aged 24 to 78 years, during 1962 and 1963 (Mittman et al., 1965). Although analysis of cross-sectional data showed a significant negative regression of chest-wall compliance on age, the regression of pulmonary compliance on age was not significant. In view of the discomfort to the subject, the large investment of time required to perform the test, and the lack of a significant age regression for pulmonary compliance, the test was not repeated after 1963.

5. Smoking History

A detailed smoking history is obtained on the first visit and is updated at each subsequent visit.

6. Chest X-Ray

A standard roentgenogram is made for postero-anterior (P-A) and lateral views of the chest with the subject standing. The total equivalent radiation exposure received by the subject above the waist is 2 rads for the two tests. Gonadal exposure is minimized by standard techniques of collimation and shielding. The techniques and equipment used are monitored by the Maryland State Health Department and the Baltimore City Hospitals' Radiation Safety Officer. X-rays were initially repeated at each visit, but since 1979 have been routinely repeated no more frequently than every five years unless there is a specific clinical indication that more frequent examination is desirable.

CARDIOVASCULAR FUNCTION

1. Screening for Cardiovascular Disease

Cardiovascular history and physical examination. At each visit, in addition to the general history and physical examination, a cardiovascular history and examination are performed by a physician of the GRC Cardiovascular Section, and a 12-lead resting electrocardiogram (ECG) is obtained. Participants who do not demonstrate definite evidence of coronary artery disease (CAD) in this examination undergo exercise stress to detect the presence of occult CAD.

On the basis of the history, physical examination, and stress testing (see below), each participant is classified as showing a) no evidence of CAD or b) evidence of probable or c) definite CAD. The following point system is employed to determine the classification of each individual:

ECG	Points
Resting	
Minn. Code 1:1	2
Minn. Code 1:2	1
Minn. Code 1:3	1
Minn. Code 4:1	1
Stress	
Minn. Code 11:1	1
Angina pectoris	
Possible	0
Probable	1
Definite	2
Previous myocardial infarction (MI)	
Possible	0
Probable	1
Definite	2

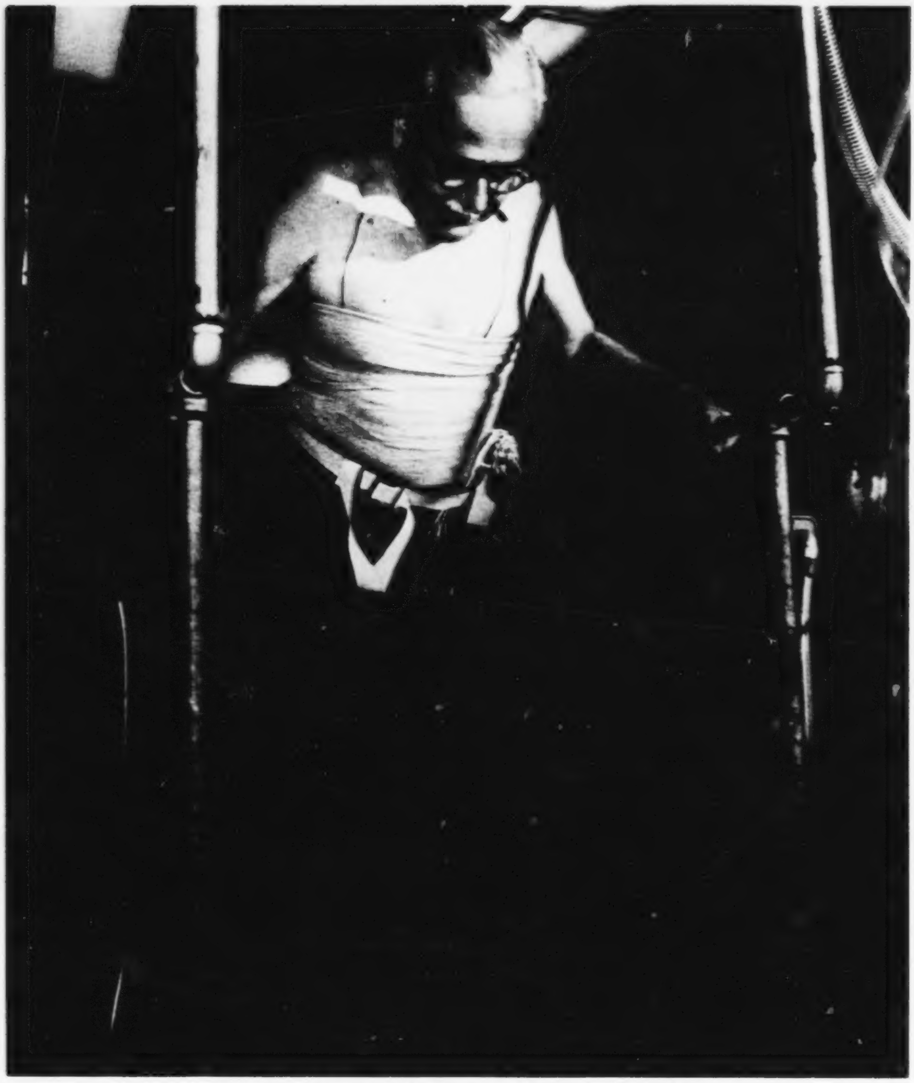


Figure IV.3. **Treadmill stress electrocardiography.**

The presence of angina and/or the history of MI is assessed by a staff cardiologist. Confirmation of prior MI also requires unequivocal hospital records. Assignment to the category of definite CAD requires at least 2 points; to probable CAD, 1 point; and to CAD-free, 0 points. The weight to be attached to the stress thallium scan in classifying CAD is presently under consideration.

Exercise stress test—double Master test. The double Master test was administered to all subjects, except those in whom a clinical contraindication was found, on all visits between 1958 and 1968. The subject climbed two nine-inch steps (total lift = 18 in or 0.46 m) so arranged that at a rate determined by his age and weight he went up two steps, went down two, turned, and went back the other way. The ECG was monitored (Master and Oppenheimer, 1929; Master et al., 1944).

Treadmill stress ECG. In 1968, a treadmill exercise test (Fig. 3), in which the subject walks on a motor-driven treadmill at a rate of 3.5 miles per hour, was introduced to replace the double Master test. At the end of each two minutes of walking, the grade of the treadmill is increased by 3% (Balke and Ware, 1959). The increments in grade are continued until excessive dyspnea or other end points (e.g., fatigue, leg pain, angina, or certain ECG abnormalities) are reached (Blomqvist, 1971). The ECG is recorded and monitored by a physician both during exercise and for six minutes afterward with the subject in the sitting position. Heart rate is estimated from the ECG. Blood pressure is measured at two-minute intervals after exercise; since 1976, it has also been measured during each stage of exercise. All ECGs are evaluated by the Minnesota Code (Rose and Blackburn, 1968).

Stress thallium scans. Since November 1977, myocardial imaging during exercise has been employed in all subjects above the age of 40 years who consent to the test. The limitation to persons over 40 was imposed because of the exposure to radiation. The thallium scan is a relatively new non-invasive technique designed to assess abnormalities in myocardial perfusion at rest and during exercise. A small amount of ^{201}thallium is injected intravenously during the final minute of maximal treadmill exercise while simultaneous multi-lead ECGs are recorded. Myocardial scans in multiple projections are made with a gamma camera linked to a computer-based display and quantification system. Redistribution scans are made two hours after the conclusion of exercise. The records, after having been read subjectively by two independent observers, are maintained for later computer processing and objective quantification. The studies were initiated in collaboration with The Johns Hopkins University and are performed at the Johns Hopkins University Hospital (Lakatta, 1978).

2. Evaluation of the Effect of Age on Cardiovascular Function

In those individuals who show no evidence of cardiovascular disease, many aspects of cardiovascular function are evaluated at rest and during stress.

Systolic time intervals. A number of time intervals during the cardiac cycle have been related to myocardial performance (Weissler et al., 1969). Simultaneous recording of the ECG, phonocardiogram, and carotid pulse made it possible to determine the duration of electro-mechanical systole, left-ventricular ejection time, and the pre-ejection period. The protocol was discontinued in 1972 with the advent of echocardiography.

Echocardiography. The echocardiogram, devised to estimate cardiac performance, records the reflection of ultrasonic waves directed toward the heart from an external probe. The record obtained by this non-invasive technique permits measurements of left-ventricular wall thickness and cavity dimension, filling and ejecting rates, ejection fraction, and left-atrial and aortic-root diameters. One-dimensional echocardiograms are recorded with the subject either supine or rotated into the left lateral position. It is anticipated that echocardiographic testing will be repeated at seven-year intervals.

Since January 1978, echocardiograms with simultaneous recording from two probes, to permit greater accuracy in the calculation of volumes of the heart chambers and detection of regional dysfunction, have been recorded in randomly selected subjects. In addition to measurements at rest, observations are also made during standardized semi-supine bicycle exercise. A semi-automated computer-assisted system quantifies structural and functional features of the left ventricle at rest and during exercise (Van Tosh et al., 1980). Measurements include myocardial and chamber areas and derived indices of left-ventricular mass and volume, mean and maximal velocity of fiber shortening, rates of diastolic lengthening, time intervals within the cardiac cycle, and regional indices of myocardial function, including left-ventricular thinning and thickening rates. This test is administered to all participants who volunteer and is supervised by the staff of the Division of Cardiology of the Department of Medicine, The Johns Hopkins University, in collaboration with the GRC Cardiovascular Section.

Multi-gated cardiac blood-pool scans (MUGA). Gated radionuclide angiography is a non-invasive method that has provided accurate and reliable estimates of left-ventricular ejection fractions both at rest and during exercise. A recent modification of the technique permits accurate measurements of left-ventricular volume throughout the cardiac cycle. It is thus an ideal technique for examination of left-ventricular function in normal human volunteers.

The protocol for the gated cardiac blood-pool scan is as follows: 2 cc of "cold pyrophosphate" are injected intravenously; after 20 minutes, 12 millicuries of technetium 99m per square meter of body-surface area are injected and supine resting gated cardiac blood-pool scan is performed. The subject then engages in graded upright bicycle exercise. Hematocrit is determined before and after maximal exercise; after exercise, a 10-cc blood sample is drawn and counted by the camera and a static marker image obtained to determine the distance between the center of the left ventricle and the chest wall. Computer analysis of the acquired images calculates left-ventricular ejection fraction, end-systolic, end-diastolic, and stroke volumes, and cardiac output for each stage of exercise. Regional wall motion is assessed visually and by computer methods (Rodeheffer et al., 1981). This test is performed at the Johns Hopkins Hospital in a collaborative study between the GRC and the Johns Hopkins Division of Cardiology.

Catecholamine secretion. At rest and during a graded treadmill exercise protocol in a subset of volunteers, oxygen consumption and plasma epinephrine, norepinephrine, and lactate were measured along with heart rate and blood pressure.

Heart-rate response to β-adrenergic stimulation. In a small subset of the population graded intravenous boluses of isoproterenol were injected in order to determine the effect of age on the heart-rate response to β-adrenergic stimulation. The end point of the protocol was an increment of 25 beats per minute over baseline.

Ventricular response to afterload stress. The left-ventricular response of young adult and old participants to hemodynamic stress was compared. Echocardiographic measurements of left-ventricular end-diastolic dimension, left-ventricular end-systolic dimension, and velocity of circumferential fiber shortening were made at rest and during 30-mm Hg increases in systolic blood pressure induced by handgrip exercise or phenylephrine infusion. In order to eliminate the influence of β-adrenergic drive, the measurements were repeated during β-adrenergic blockade with propranolol. The efficacy of the block was tested by demonstration that isoproterenol infusion did not result in an increase in heart rate (Yin et al., 1978).

Non-invasive His-bundle electrocardiography. Previous studies have shown an age-related increase in atrioventricular (AV) conduction time (PR interval) but have not localized its origin in relation to the His bundle. A microprocessor-assisted high-resolution ECG has been employed on selected subjects since January 1981. This technique uses signal averaging and amplification of 512 cardiac cycles as well as filtration of random noise to record low-amplitude ECG signals from the body surface. From these recordings, a His-bundle potential can be identified in a substantial percentage of individuals. This allows an assessment of AV conduction both proximal to the His bundle (PH interval) and distal to it (HV interval) (Das et al., 1982).

Inotropic effect of digitalis. Because digitalis is commonly prescribed in the elderly and is attended by a high risk of toxicity, averaging 20% in most series (Beller et al., 1971), it is of vital concern whether its inotropic efficacy is decreased with advanced age in man. To answer this question, one-dimensional echocardiography and systolic time intervals have been used to measure the effect of an intravenous bolus of ouabain, a rapidly acting digitalis glycoside, on cardiac performance in healthy men. After baseline measurements are obtained, β-adrenergic blockade is induced by intravenous propranolol in order to eliminate variations in resting sympathetic tone that could obscure the effects of digitalis. Echograms and systolic time intervals are obtained periodically for 90 minutes after injection of ouabain.

Ambulatory ECG and blood-pressure monitoring. Since July 1978, an ambulatory 24-hour ECG has been recorded in normal BLSA participants during their routine activities at the GRC. The two-channel ECGs are analyzed by a semi-automated computer technique for heart rate, ectopic beats, and disturbances in conduction. Thus far 150 participants, all of whom were clinically healthy and had normal ECG responses to maximal treadmill exercise, have been studied by this technique.

Since January 1981, ambulatory ECGs and ambulatory blood pressure have been recorded automatically over a 24-hour period in normal subjects of all ages. Blood pressure is taken automatically every 7½ minutes and recorded on the same tape used to record the ambulatory ECG. This technique allows assessment of diurnal variability and absolute level of blood pressure as a function of age.

Peripheral blood flow. Venous-occlusion plethysmography was initiated in July 1980 to assess aging changes in maximum peripheral blood flow to the leg. By this technique the leg is placed in a water-filled box equipped to measure changes in limb volume. A blood-pressure cuff placed around the leg proximal to the plethysmograph is used to occlude arterial flow for periods of one to five minutes. The cuff pressure is then rapidly lowered to a level sufficient to occlude venous outflow but not arterial inflow. The consequent increase in leg volume, which is equivalent to arterial inflow, is then calculated from the rate of change in water pressure in the box with time. The procedure is initially performed at a water temperature of 27°C; it is repeated at 32°C to assess the increase in arterial flow generated by thermal stress. Such measurements allow assessment of maximal arterial flow and recovery rates from the induced limb ischemia. These values may then be correlated with known risk factors and predictors for the development of peripheral vascular disease.

CARBOHYDRATE METABOLISM: THE GLUCOSE-INSULIN HOMEOSTATIC SYSTEM

Tests to evaluate the glucose-insulin homeostatic system were given during the following periods:

Intravenous glucose tolerance test (IVGTT), January 1963-June 1977.
Intravenous insulin tolerance test (IVITT), June 1963-June 1964.

Cortisone glucose tolerance test (CGTT), January 1964-June 1977.

Oral glucose tolerance test (OGTT), 1.75 g/kg body weight, July 1964-June 1977; since June 1977 the glucose dose has been 40 g/m^2 surface area.

Intravenous tolbutamide response test (TRT), January 1965-June 1977.

One of the tests is usually performed on each visit. The four tests that were administered until June 1977 were generally given in a series that was repeated after all four had been administered. Since that date only the OGTT has been done. All are performed under basal conditions.

Activity prior to the test is limited, since subjects spend the preceding night at the Center as part of their 2½-day stay. Subjects remain in bed, in a reclining or a semi-reclining position, during the test. Smoking is not permitted before or during the test. Fasting and all subsequent venous-blood samples are obtained through an indwelling catheter.

Glucose was determined by a manual glucose oxidase method until September 1963; by the AutoAnalyzer automated ferricyanide reduction method (Technicon Instruments Corp., Tarrytown, New York) until June 1977; and thereafter by an automated glucose oxidase method (Beckman Instruments, Inc., Fullerton, California). Initially, from 1963 to 1966, whole-blood samples were deproteinized by the Somogyi technique (1945). Since 1966, plasma samples have been analyzed without deproteinization. Factors for conversion from the manual glucose oxidase to the automated ferricyanide method and from whole-blood to plasma samples were determined by simultaneous analyses of multiple specimens by the older and the newer methods. Since the AutoAnalyzer and Beckman methods gave nearly identical results, it was not necessary to apply a conversion factor.

The tests have been performed as follows:

Intravenous glucose tolerance test (IVGTT). A 20% solution of dextrose in water (0.375 g/kg body weight) was infused at a constant rate over a five-minute period. Blood samples were collected every ten minutes until 60 minutes; in the earliest studies, a final sample was collected at 80 minutes.

Intravenous insulin tolerance test (IVITT). A dose of 0.05 units of crystalline zinc insulin per kg body weight was injected nearly instantaneously. Venous-blood samples were collected from an indwelling catheter at ten-minute intervals for one hour.

Cortisone glucose tolerance test (CGTT). Cortisone acetate was given by mouth 8.5 and two hours before glucose ingestion, in two equal doses determined by body weight ($<$ 124 lb = 37.5 mg; 124–159 lb = 50 mg; 160–194 lb = 62.5 mg; 195–230 lb = 75 mg; and $>$ 230 lb = 87.5 mg). The glucose dose was 1.75 g/kg actual (as opposed to "desirable") body weight, given as a 30% solution flavored with lemon juice, which was ingested in ten minutes or less. Blood samples were drawn every 20 minutes for two hours (Pozefsky et al., 1965).

Oral glucose tolerance test (OGTT). The technique was originally the same as that used for the CGTT, but the steroid administration was omitted. In July 1977 the glucose dose was changed to 40 g/m^2 surface area, in accordance with the recommendation of the Committee on Statistics of the American Diabetes Association (Klimt et al., 1969).

In 1979 the National Diabetes Data Group recommended that 75g of glucose be given to all subjects regardless of body size. This dose is equivalent to 40 g/m^2 surface area for an average-sized person. We have elected to continue use of the 40 g/m^2

dosage, which the Data Group had rejected as impractical for the usual clinical test situations. Considering a dose adjustment for body size an advantageous detail, we decided not to introduce still another technical change into our study.

Intravenous tolbutamide response test (TRT). One gram of sodium tolbutamide in a 5% solution per 70 kg body weight was injected through an indwelling intravenous catheter in two minutes. Zero time was taken as the midpoint of the injection. Blood samples were obtained at 2, 6, 10, 15, 20, 30, 45, and 60 minutes.

Performance was judged primarily by the percentage of fall in glucose concentration at 20 and 30 minutes (Swerdloff et al., 1967).

Glucose-clamp test. A manual feedback technique, the glucose clamp, makes it possible to maintain blood glucose at any level chosen by the investigator (Andres et al., 1966; DeFronzo et al., 1979). Two basic types of study have been performed. In the "euglycemic clamp study," insulin is infused at a constant rate while the blood-glucose concentration is maintained at the subject's basal level. This is primarily a test of sensitivity of body tissues to insulin. In the "hyperglycemic clamp study," the plasma-glucose concentration is raised rapidly to a hyperglycemic plateau and is maintained at that level for two hours. The glucose plateaus studied have been 54, 98, 143, or 231 mg/dl above the basal level. This is a test of pancreatic beta-cell sensitivity to glucose and sensitivity of body tissues to insulin. These tests were given to subjects selected for special characteristics with respect to glucose tolerance and obesity. The schedule for retesting has not yet been determined.

METABOLISM OF DRUGS

1. Antipyrine Metabolism

Antipyrine, which is almost entirely metabolized by the liver, has been used as a marker drug for the hepatic microsomal enzyme system. A cross-sectional study of antipyrine metabolism was carried out between 1958 and 1967. Healthy subjects who were receiving no potentially interfering medication were included in the analysis of the interactive effects of age, smoking, consumption of caffeine, and alcohol intake. The subjects received an intravenous infusion of antipyrine (1 g/30 ml isotonic saline) in a 20-minute period. Plasma levels of antipyrine were measured at 0, 2, 4, 6, and 8 hours by the method of Brodie et al. (1949). The overall elimination rate constant (k_e) and the theoretical plasma concentration at zero time (C_0) for each subject were determined from regression analysis of the natural log of the plasma concentration with respect to time. Biologic half-life, apparent volume of distribution, and metabolic clearance rate were computed on the assumption of a single distribution volume and simple exponential decay by standard pharmacokinetic formulae (Vestal et al., 1975).

2. Ethanol Metabolism

In 1974–1975, a cross-sectional multidisciplinary study of ethanol metabolism and aging was undertaken (Vestal et al., 1977). Observations were made not only of ethanol kinetics, but also of plasma arginine vasopressin (AVP) response and reaction time and memory. The subjects, who had abstained from alcohol for three weeks, received a one-hour infusion of ethanol at a rate of 375 mg/m^2 surface area per minute via an antecubital intravenous catheter. Blood samples for measurement of blood-ethanol concentrations were obtained at frequent intervals during the infusion and for four hours post infusion. Blood ethanol was assayed by a modification of the methods of Payne et al. (1967) and Roach and Creaven (1968). Ethanol kinetics were computed by the compartmental analysis of Berman and Weiss (1976). Body composition was calculated by the anthropometric method of Behnke (1961).

The effect of ethanol on plasma AVP levels (antidiuretic hormone) was assessed by radioimmunoassay (Robertson et al., 1973). Reaction-time and memory tests were carried out during and after the ethanol infusion by techniques similar to those described below under "Learning, Memory, and Decision Tasks."

HYPOTHALAMIC-PITUITARY FUNCTION

1. Reproductive Hormone System

This research, which was performed from 1977 to 1979, evaluated the hypothalamic-pituitary-testicular system, then correlated the endocrine studies with sexual history. Baseline plasma samples were obtained and a two-hour LHRH test was performed to measure pituitary gonadotropin reverse. This was followed by injection of human chorionic gonadotropin (hCG) to evaluate testis secretory reserve. A second injection was given the next morning. Blood samples were obtained after the injection of hCG and the following morning.

Plasma gonadotropins were assayed by double-antibody radioimmunoassay; plasma testosterone and dihydrotestosterone by radioimmunoassay; plasma estrone and estradiol by charcoal radioimmunoassay; and free testosterone by an ion-exchange technique (Harman and Danner, 1977; Harman and Tsitouras, 1980; Harman et al., 1980; S.M. Harman, 1981). Stored lyophilized plasma samples obtained at earlier ages were also assayed and their testosterone levels correlated with the sexual-behavior histories of the same subjects.

2. Hypothalamic-Posterior Pituitary Function

The effect of age on this endocrine system was determined cross-sectionally by assessment of the change in plasma-AVP levels in response to both a secretory stimulus (hypertonic saline) and an inhibitory stimulus (ethanol). The tests were performed from 1974 to 1976 in a small, carefully selected subset of the BLSA population.

Ethanol infusion test. In a subset of the subjects who were given ethanol intravenously, AVP responses were followed (see above, "Metabolism of Drugs," for a description of the ethanol-infusion protocol). Subjects abstained from all alcohol for at least 21 days before the study. AVP was measured by radioimmunoassay (Robertson et al., 1973). Samples of venous blood were obtained at short intervals during the ethanol infusion and for five hours thereafter (Vestal et al., 1977).

Hypertonic saline infusion test. Very careful screening of volunteers to exclude subjects with renal, hepatic, or cardiac disease was necessary. A two-hour infusion of 3% NaCl was given at a rate of 0.1 ml/kg body weight per minute after a 12-hour period of dehydration. Blood samples were collected every 20 minutes during the infusion (Helderman et al., 1978).

THE IMMUNE SYSTEM

Assessment of immune function is centered on the determination of serum antibody levels, serum immuno-protein levels, the function of peripheral blood lymphocytes *in vitro*, and the analysis of granulocytic cell function, also *in vitro* (tissue culture). The assays measure immune function in separate areas: the thymic-dependent area (cell-mediated immunity), the bone-marrow-dependent area (humoral immunity), and the area of nonspecific host resistance to infective organisms.

The several assays for cell-mediated immune function include proliferative response to a mitogenic agent, the ability to cooperate with antibody-forming cell-precursors to initiate an immune response, the enumeration of T lymphocytes and T-cell subset populations using morphologic criteria, and the measurement of lymphocytic ability to kill tumor cells.

The assays for the "B"-cell activity (humoral immunity), which include measurements of cellular activity, serum-protein concentrations, and morphologic identification, are based on the functional ability of some lymphoid cells to make antibodies. The assay for granulocytic cell activity measures metabolic activity during a period in which the cells are phagocytosing latex particles, as well as the ability of granulocytes to kill phagocytized bacteria.

The tests are carried out on leucocytes separated from fasting blood samples drawn from all subjects. The program, initiated in 1978, anticipates repetition of tests at six-year intervals (Adler et al., 1977).

CELL REPLICATION

Since previous studies had shown that the number of replications of human cells grown in culture is inversely related to the age of the donor (Martin et al., 1970; Schneider and Mitsui, 1976), a study of *in-vivo* human cellular aging in skin fibroblast cultures was introduced into the BLSA in 1974. Skin fibroblast cultures are established from a 2-mm punch biopsy obtained from the inner aspect of the left upper arm from male volunteers aged 17 to 96 years. In addition, biopsies are repeated at intervals of from three to five years after the initial sampling.

Cells are cultured under standardized conditions (Schneider and Mitsui, 1976). Initially, several observations were made on successive transfers of each culture: time of onset of senescent phase (failure of culture to reach confluency within one month of transfer), *in-vitro* life span of culture, cell-population replication rate, percentage of replicating cells in the culture, cell number at confluency, percentage of cells able to form large colonies, receptors for insulin and epidermal growth factor, prostacyclin synthesis, viral replication, and sister chromatid exchanges per cell. Subsequent observations have been more limited in their scope and include only culture life span, time of onset of cell senescence, and capacity for colony formation.

Cultures derived from each subject are frozen and stored at the GRC for direct comparison with other cultures taken from the same subject at a later visit. In addition, subcultures from 100 biopsies have been forwarded to the Aging Cell Repository, Institute for Medical Research, Camden, New Jersey, where they are now available for study by interested investigators.

SPECIAL SENSES

1. Eye Tonography

This test is administered to all subjects on each visit. A four-minute Schiötz test (Drews, 1967) is performed for each eye; intraocular pressure is increased by a 5.5-g pressure transducer placed on the cornea, which has been locally anesthetized by tropicamide, and the change in pressure is recorded over a four-minute period. Subjects with histories of corneal injuries are not tested.

2. Visual Screening Tests

Since 1964, the Titmus Optical Vision Tester has been used to evaluate visual function in all subjects (Titmus Optical Co., Petersburg, Virginia, 1959). The following observations are recorded: acuity (binocular, right eye and left eye, near and far); stereopsis (depth vision); color discrimination; vertical phoria; and lateral phoria, near and far. These tests were administered to each subject on alternate visits from 1964 to 1974; since 1974, they have been administered at each visit.

3. Fundus Photography

Since 1975, retinal photographs have been taken to identify vascular changes in the eye. Stereoscopic views of the macula and the optic-disc areas of the fundus of each eye are photographed in color with the Zeiss Fundus Camera. Since the procedure requires dilation of the pupils, subjects with histories of angle-closure glaucoma or Schiötz pressures (see above, "Eye Tonography") of 23 mm Hg or greater are not tested. This test is conducted at the GRC in collaboration with the staff of the Wilmer Eye Institute, The Johns Hopkins University.

4. Audiometry

Pure and pulsed-tone audiometric tests are performed with the Bekesy Audiometer, Model E800 (Bekesy, 1947; Corso, 1955; Hirsh, 1962). The test is administered with the subject seated in a soundproof cabinet. Tones are presented to each ear through air-conduction headphones. The subject's task is to depress a switch when he hears the tone and to release the switch when the tone disappears. The switch controls the motor-driven attenuator of the audiometer: When it is depressed, the signal intensity decreases; when it is released, signal intensity increases. A pen connected to the attenuator traces a continuous record of the subject's intensity adjustments on an audiogram form, producing a graphic representation of his threshold. Auditory thresholds are determined at pure tone frequencies between 150 and 8000 Hz.

Between 1965 and 1973 this test was administered to each subject on alternate visits; since 1974, it has been administered on every visit.

COGNITIVE PERFORMANCE

1. Intelligence Tests

The Army Alpha examination (forms A and B) is a combination of the five forms of Alpha used in the United States Army during World War I. The questions best for general use were selected by order-of-merit method. Items addressed specifically to male recruits were excluded and military terms modified (Bregman, 1925, 1947).

Form A of this paper-and-pencil examination, consisting of eight subtests, has been administered to all subjects since 1960. Form B is administered on visit 5 (six to eight years later), and Form A is repeated on visit 9; the two forms are given alternately on all subsequent visits. When the subject reaches age 70 and is tested every year, the form of the test given on the last previous visit is repeated. A speed score is obtained for each subtest by stopping the subject after the time specified, while a power score is obtained by permitting the subject to spend as much time as he likes in completing each test.

The Vocabulary Test (WAIS). This subtest from the Wechsler Adult Intelligence Scale (Wechsler, 1955) has been administered at six-year intervals since 1960. The task is to define 40 words.

Southern California Tests of Mental Ability. From 1959 to 1978 this battery of tests (Christensen et al., 1958; Christensen and Guilford, 1959) was administered twice, at a six-year interval. First-time testing was discontinued in 1972; by 1978, all participants who had taken the earlier test had completed it a second time, and the test was discontinued. The five timed tests (Associational Fluency, Expressional Fluency, Ideational Fluency, Word Fluency, and Consequences) were designed to measure several aspects of creative thinking. For example, in Expressional Fluency the task is to write meaningful four-word sentences in which the initial letter of each word has been specified; in Ideational Fluency, the subject lists items that meet such specific criteria as "fluids that burn."

2. Learning, Memory, and Decision Tasks

Verbal learning. Data collection began in 1960 for two studies of verbal learning. In serial learning, a list of familiar words is presented repeatedly in the same order. The task is to respond to each word with the next word in the list. In paired-associate learning, items that consist of a stimulus (two consonants) and a response component (a familiar adjective) are presented repeatedly in different orders. The task is to say the word that is paired with the two consonants. For both tasks, each subject was assigned to one of three pace conditions determined by the amount of time permitted for each response. Total errors and trials to criterion are the measures (Arenberg, 1967b; Arenberg and Robertson-Tchabo, 1977). These tasks are repeated at six-year intervals with different sets of words and consonants.

Benton Visual Retention Test. This non-verbal memory test (Benton, 1963) has been administered at six-year intervals since 1960. Form C is used for the first administration, Form E for the second, Form D for the third, and Form C again for the fourth. Each form is made up of ten designs with one or more figures; the task is to reproduce each design from memory after inspecting it for ten seconds. The primary measure is the number of errors in all ten reproductions *(Arenberg, 1978).*

Memory and decision tasks. In 1978 the following set of memory and decision tasks was introduced:

- Single-trial, immediate free recall (IMFR). Each of four lists consists of 12 familiar nouns. After the words are shown paced, the task is to report as many words as possible.
- Forward digit memory. The task is to recall, in order, lists of three to nine digits presented auditorily.
- Delayed memory. After each IMFR list and an interpolated task (forward

digit memory), one of two delayed memory procedures is administered. In delayed free recall, the task is to report as many of the words as possible from the previous IMFR list. In delayed recognition, the 12 words from the previous IMFR list and 12 distractor words are shown one at a time, and the task is to decide whether the word has already been presented.
- Dichotic listening. The task is to identify two digits presented simultaneously, one to each ear. Each set consists of 28 pairs.
- Decision tasks. These tasks require response to the visual presentation of designated digits under five different conditions. With the exception of the first task, the display is paced at a rate of one digit per second. The first task is to respond to the onset of a zero. The second is to respond to a specified digit. The third is to respond to any even or odd digit. The fourth is to respond to an even-odd or odd-even sequence of digits, and the fifth to respond to any two consecutive even or consecutive odd digits. Decision time and accuracy are the measures.

The same set of tasks, with different word lists, is to be repeated six years after the first administration.

3. Problem Solving

Logical problem solving. An experimental procedure was designed to measure effectiveness of reasoning. The apparatus consists of a display with six numbered and three lettered lights, each of which has an adjacent push-button, and a central light (G) that has no button. Each problem contains a set of logical relations indicated by arrows between lights. The ultimate task in each problem is to arrive at the outcome, G, via a sequence of inputs. The number of uninformative inputs is the primary measure. A set of logically identical problems is administered at least six years later. From 1962 to 1966, each problem was presented as a single task *(Arenberg, 1974)*. From 1966 to 1974, for subjects who were administered these problems for the first time, each problem was presented in two parts to obtain independent measures of performance in analysis and synthesis.

Concept identification. The ability to identify concepts in the context of a problem-solving task is also evaluated. Each of 12 concept problems requires the identification of one or two "poisoned" foods. The subject selects "meals" consisting of four of the foods on a list, and the experimenter indicates whether that "meal" is fatal. The task is to identify the "poisoned" foods with as few "meals" as possible. The two primary performance measures are correctness of the identification and effectiveness in reaching a solution as indicated by the number of "meal" selections. Concept problem solving, initiated in 1967, is administered at six-year intervals.

PERSONALITY AND DEVELOPMENTAL CHARACTERISTICS

1. Personality

Guilford-Zimmerman Temperament Survey. A questionnaire consisting of 300 items provides an assessment of ten traits: General Activity, Restraint, Ascendance, Sociability, Emotional Stability, Objectivity, Friendliness, Thoughtfulness, Personal Relations, and Masculinity (Guilford and Zimmerman, 1956; Guilford et al., 1976). Each subject is given the standardized instructions individually and completes the questionnaire during the remainder of his visit to the GRC. Until 1978, the test was administered every six years; since that time the interval has been 12 years.

Eysenck Personality Inventory. A standard measure of personality, yielding scores for Neuroticism and Extraversion, as well as a Lie scale. There are 57 items in a yes-no format.

NEO Inventory (Costa and McCrae, 1980c). A 145-item personality questionnaire that measures six traits in each of three broad domains of personality: Neuroticism, Extraversion, and Openness to Experience. Questions are answered on a five-point Likert scale: "strongly disagree," "disagree," "neutral," "agree," "strongly agree."

NEO Rating Inventory (McCrae, 1982b). Spouses of BLSA participants are asked to complete the NEO Rating Inventory. Measures of six traits in each of three broad domains of personality are obtained to provide an alternative to self-report measurement of personality (see above, "NEO Inventory").

Perceptual tests. In a one-to-two-hour session at the GRC, subjects are administered three psychological tests: the Thematic Apperception Test (TAT), in which subjects are shown a series of pictures and asked to tell a story about each; an Embedded Figures Test, in which subjects locate a hidden figure in a complex design; and the Holtzman Inkblot Test, in which subjects are presented with a series of inkblots and asked to tell what they call to mind. Ninety-six participants have taken these perceptual tests once since 1979.

Social desirability scale (Crowne and Marlowe, 1964). Thirty-three yes-no questions are asked to measure the subject's tendency to give socially acceptable answers. The scale has been used as a measure of the need for approval as well as of defensiveness.

Imaginal Processes Inventory—daydreaming. Aspects of daydreaming and related imaginal processes are determined for all participants by their responses to the 344-item Imaginal Processes Inventory developed by Singer and Antrobus (1963, 1972; Singer, 1975), as revised in 1970. The inventory is usually administered to single subjects or small groups. Each participant is given a brief definition of daydreaming and an explanation of the general purposes of the study. Completion of the inventory is self-paced and without supervision. Each item has five response options representing points on a continuum implying frequency or quality. A total of 28 scales are determined from responses to non-overlapping items. Each scale contains 12 items, except one that contains 20. The scales measure the content and structure of daydreaming by items drawn from intensive interviews; items are both specific (e.g., "I daydream about saving a drowning child") and general (e.g., "In my daydreams I feel guilty for having escaped punishment") (Giambra, 1977b, 1977–78). The scales have internal consistency and test-retest reliability (Giambra, 1974). The inventory was introduced into the BLSA in 1972–1973 and is repeated at six-year intervals.

2. Developmental Antecedents

Parent-child relations questionnaire (Roe and Siegelman, 1963). Subjects record their perceptions of relations with their parents when they were children. Fifty questions yield scores for casual-demanding, love-rejection, and attention dimensions. There are separate forms for son-mother, son-father, daughter-mother, and daughter-father.

Activities and attitudes questionnaire. The schedule and inventory entitled "Your

Activities and Attitudes" (Cavan et al., 1949) is given to each participant to be filled out without supervision during his first visit to the GRC and readministered at every fourth visit. The inventory is composed of three parts: background information, including general information about the participant and his earlier life; an activity inventory; and an attitude inventory.

The activity inventory provides eleven subscores in such areas as leisure-time and religious activities, intimate personal contacts, security, and health status. The attitude inventory deals with the personal aspects of adjustment. It contains eight groups of statements concerning health, friends, work, economic security, religion, and feelings of usefulness, happiness, and family.

STRESS AND COPING PROCESSES

1. Stress

Schedule of Life Events (SLE). This is a checklist, completed each two years, of recent potentially stressful events. Subjects rate their perceived stressfulness.

Daily events checklist. Subjects are asked to indicate which of a series of minor daily stresses and strains they have recently experienced, and to rate their pleasantness or unpleasantness.

Stress-and-coping interview. A 90-minute private interview is given by a psychologist or psychiatrist concerning history of stresses and coping efforts. The interview is videotaped, and standard ratings are made by the interviewer and another rater.

2. Coping and Defense Mechanisms

Coping self-interview. Participants are asked to nominate three events they have recently experienced: a challenge, a threat, and a loss. For each, they are asked if they used any of a set of 50 coping responses in dealing with the problem they have selected. In addition, they indicate whether the response helped to solve the problem or made them feel better.

Coping questionnaire. From among recent life events experienced by subjects, one target event is selected. Participants are asked to indicate which of 118 ways of coping they used in dealing with the event. Scores for 28 different coping mechanisms are derived.

Defense-mechanism inventory (Gleser and Ihilevich, 1969). A series of stories is presented, and subjects are asked to imagine how they would respond in thought, actions, feelings, and fantasy to the circumstances described. Five dimensions of defensive processes can be scored from the instrument.

3. Adaptational Outcomes

Well-being assessment sheet. This instrument, administered each two years, assesses psychological well-being, satisfaction with various areas of life, and overall evaluation of life.

Profile of mood states. Subjects indicate on this form the level of disturbance in seven moods: tension, anger, depression, fatigue, vigor, friendliness, and confusion. There is also a total mood score. Different forms allow for administration under "right now" or "in the past week" conditions.

MARITAL AND SEXUAL EXPERIENCE

Since 1967, interviews have been conducted with BLSA males concerning their current and past experience of marriage and sexual activity. Data are collected in a single two-hour session by a sociologist with extensive experience in such interviews (Martin, 1975). Over the years the refusal rate has varied between 2% and 3%. In all, 777 men have completed interviews.

Each interview follows a predetermined series of questions that have been memorized by the investigator; information is recorded in a code. Subjects are asked about the presence or absence of coitus, masturbation, nocturnal emission, and homosexual activity in their adult lives, the age of the subject at onset of each, and the frequency of their occurrence in relation to age and marital status. Since these behaviors account for nearly all male orgasmic experience (Kinsey et al., 1948), their combined frequencies constitute a measure of sexual functioning that is essentially unobtainable by other means. These frequencies, expressed in the interview as times per week, per month, or per year, are then converted into the number of sexual events falling into each five-year interval between age 20 and the time of report. Additional questions elicit information on sexual attitudes and reactions and characterize other aspects of the participants' marital, residential, religious, occupational, educational, military, and parental-home experience. None of the interviews has been repeated.

CHAPTER V

Cross-Sectional Studies of Aging in Men

The subjects from the Baltimore Longitudinal Study of Aging (BLSA) provide a rich resource for cross-sectional studies to extend knowledge about age differences in normal successful men living independent lives in the community. This chapter summarizes the results of many of the cross-sectional studies, which define average differences between groups or the average regression of the variable on age. To identify outcomes or the effects of specific events on later performance, the subjects must be re-examined or information about them must be gathered at a later time. Although longitudinal observations are helpful in identifying time sequences, they are not essential in determining outcomes.

Longitudinal studies in which serial changes are based on the analysis of repeated observations in the same subjects are described and summarized in Chapter VI, whose final section also summarizes studies of outcomes based on a single characteristic, such as survival.

Some of the studies reported in this chapter are based on tests that were systematically repeated, so that longitudinal analyses were ultimately possible. Others describe results of tests not repeated for a variety of reasons, among which were: The average trend with age was so small in comparison with the variance among subjects that the age regression lacked statistical significance; the test required more time with the subject than could be provided within the testing schedule; analytical procedures required more laboratory assistance than was available; the primary investigator had left the Gerontology Research Center (GRC); or from the outset there was no requirement that the test be repeated, as was the case with the study of ethanol metabolism (Vestal et al., 1977) described below.

In some instances it was possible to carry out well-designed interdisciplinary studies because of the close association among investigators from different scientific disciplines, who were brought together primarily because the BLSA provided a well-characterized population of normal males. Again, a good example is the detailed study of the effects of age on the physiological responses to ethanol, which required close collaboration among physicians, physiologists, pharmacologists, endocrinologists, and psychologists.

PHYSIOLOGY

1. Cell-Culture Senescence and *In-Vitro* Life Span

Although differences observed at the organelle and macromolecular levels in early and late passage cell cultures have been attributed to cellular "aging," there is concern that such changes may not accurately reflect human cellular aging *in vivo*. This problem was addressed in a study (Schneider and Mitsui, 1976) designed to determine: a) whether differences would be observed in the onset of cell-culture senescence and in the cumulative replication capacity of fibroblast cultures derived from 2-mm skin-punch

Table V.1. Characteristics of Skin Explants and Cell Cultures Derived from Young and Old Human Donors

	Young Donors (21–36 yr)	Old Donors (63–92 yr)
Successful cell outgrowth from explants	23/34	24/39
Cell cultures that senesced before 10 CPD[c]	0/23	3/24
Explant outgrowth at 1 week (units)	2.44 ± 0.30 (26)[a,b]	1.44 ± 0.15 (29)[b]
Onset of senescent phase (*A*) (CPD)	35.2 ± 2.1 (23)	20.0 ± 2.0 (24) 22.5 ± 1.7 (21)[d]
Onset of senescent phase (*B*) (CPD)	41.6 ± 2.4 (23)	26.3 ± 2.6 (24) 29.6 ± 2.1 (21)[d]
In-vitro lifespan (CPD)	44.6 ± 2.5 (23)	29.8 ± 2.9 (24) 33.6 ± 2.2 (21)[d]
In-vitro lifespan (days)	273 ± 11 (23)	218 ± 14 (24) 236 ± 12 (21)[d]

[a] Values are expressed as mean ± standard error of the mean. Numbers within parentheses indicate the number of individual cell cultures examined.

[b] Includes several measurements of outgrowth from two explants taken from the same donor.

[c] CPD = cell-population doublings.

[d] If only cultures with > 10 CPD included.

From Schneider and Mitsui (1976).

biopsies taken from young and old BLSA subjects; and b) whether parameters that change with increased *in-vitro* "aging" are altered as a function of donor age.

In comparison with fibroblast cultures derived from young donors, those derived from elderly donors showed statistically significant decreases in migration, *in-vitro* life span, cell-population doubling (CPD) rate, and cell number at confluency, although no significant differences were found in modal cell volumes or, DNA or RNA content (Tab. 1). While these findings confirm the utility of fibroblast cultures for studying human cellular aging (Martin et al., 1970), the differences observed in cell cultures derived from young and old donors varied both quantitatively and qualitatively from the *in-vitro* "aging" seen in early and late passage WI-38 cells. Also, several changes that occurred *in vitro* were not a function of donor age. It was concluded that early and late passage human diploid cell cultures may provide a useful system for examining loss of replicative potential, but that fibroblast cultures derived from old and young donors may be more appropriate for the study of human cellular aging. The continuing participation of the fibroblast donors in the BLSA presents an opportunity for follow-up studies to determine how various *in-vitro* indices correlate with physiologic studies and whether *in-vitro* life span has any relation to *in-vivo* life span or to the development of age-related disorders.

This work was subsequently extended by Smith et al. (1978), who compared the growth patterns of BLSA skin fibroblasts with fetal lung fibroblasts to determine the relation between colony-size distribution and donor age. Fetal cell cultures and cultures established from skin biopsies of old (64+ yr) and young (20–34 yr) donors were examined after two weeks of incubation (5–15 population doublings). Both in human fetal lung and in adult skin fibroblast cultures the distribution of colony sizes (which ranged from one to several thousand cells) was an accurate indicator of the number of subsequent *in-vitro* population doublings that could be attained by the parent culture. In addition, the colony-size distributions were related to the chronological age of the cell-culture donor. The percentage of large colonies with significant proliferative ability was thus highest in cultures of fetal origin, intermediate in cultures from young adults, and lowest in cultures from the over-65 age group. It was concluded that colony-size distributions achieved in tissue culture are good indicators of both *in-vitro* and *in-vivo* human cellular aging.

These studies led to more extensive analyses of the relation between *in-vitro* measurements and *in-vivo* human cellular aging. The establishment of cell cultures derived from 400 participants in the BLSA by the technique of cell-banking (Schneider, 1979) will make possible such correlative analyses as the study of the relation between *in-vivo* glucose tolerance or immune responses and such *in-vitro* functions as cell replication. Perhaps the major question this study will address is whether *in-vitro* data will provide significant *in-vivo* predictive information about the original donor. It will also permit a longitudinal follow-up study of serially derived cultures from the same donors over extended periods.

Cellular aging has been described as a progressive conversion of proliferating cells from a cycling to a non-cycling state (Gelfant and Grove, 1974). A study (Tice et al., 1979) of the age-related decline in immunocompetence as it is manifest in a loss in cell-mediated immunity also found both a progressive inability of normally quiescent cell populations to respond to a proliferative stimulus and an increase in cell-cycle durations. Peripheral lymphocytes stimulated by phytohemagglutinin (PHA) were examined by the bromodeoxyuridine staining technique. Peripheral lymphocytes from aged subjects (> 75 yr) were stimulated at about one-half the rate of those from young subjects (< 21 yr). Cell-cycle durations were determined to be 10.0–25.0 hours in cultures from aged and 10.6–15.6 hours in cultures from young subjects.

Although the findings do not allow a determination whether the increase in cell-cycle durations is due to a slowing of all phases of the cell cycle or of one particular phase, the aging of cell populations capable of proliferation may perhaps be attributable to alterations in transition probability—a mathematical expression that defines the ability of a cell to initiate a proliferative response somewhere in the G_1 phase of the cell cycle.

2. Body Composition

Body composition and metabolism. Measurement of body composition is important in determining the degree to which the age-related changes in human functional capacities are due to simple loss of tissues (cells) or to reduced function in the tissue that remains. Age trends in selected indices of body composition were therefore measured in a study of 143 BLSA participants 20 to 99 years of age (Norris et al., 1963), to observe the relations among different methods of estimating body composition applied simultaneously in the same individual. Body fat was estimated by three mathematical procedures that use data on body density, body-water compartments, muscle mass, and bone-mineral mass. Measurements on each subject, made within a two-day period, included body-water spaces (total body water from antipyrine space and extracellular water from thiocyanate space), body density (helium displacement), x-ray bone density, height, weight, 24-hour creatinine excretion, and basal oxygen consumption. Significant annual decreases were found for height (–0.08%), creatinine excretion (–0.77%), bone density (–0.48%), and basal oxygen consumption (–0.41%). Subsequent longitudinal analyses of these observations are reported in Chapter VI.

Although measurement of the basal metabolic rate (BMR) has been replaced as a clinical tool by other tests of thyroid function, it is still of scientific interest as an estimate of the "active" cellular mass—the amount of functioning tissue—in studies of aging. Tzankoff and Norris (1977) examined the relation between basal oxygen consumption and 24-hour creatinine excretion in 959 healthy males aged 20 to 97 years in order to identify that component of the lean body mass responsible for the age-related decrease in the BMR. An age-independent linear relation was found between paired values of basal oxygen consumption, a measure of total metabolic activity, and 24-hour creatinine excretion, a measure of muscle mass. Although no age differences were found in mean basal oxygen consumption in subjects up to age 45, the average values were significantly lower for each decade thereafter; only the difference found in the oldest group failed to reach statistical significance (Fig. 1). Each succeeding age-group also had lower mean values for 24-hour creatinine excretion; the differences were statistically significant for all but the youngest group (Fig. 2). Since creatinine excretion was assumed to be proportional to muscle mass, the data indicated that

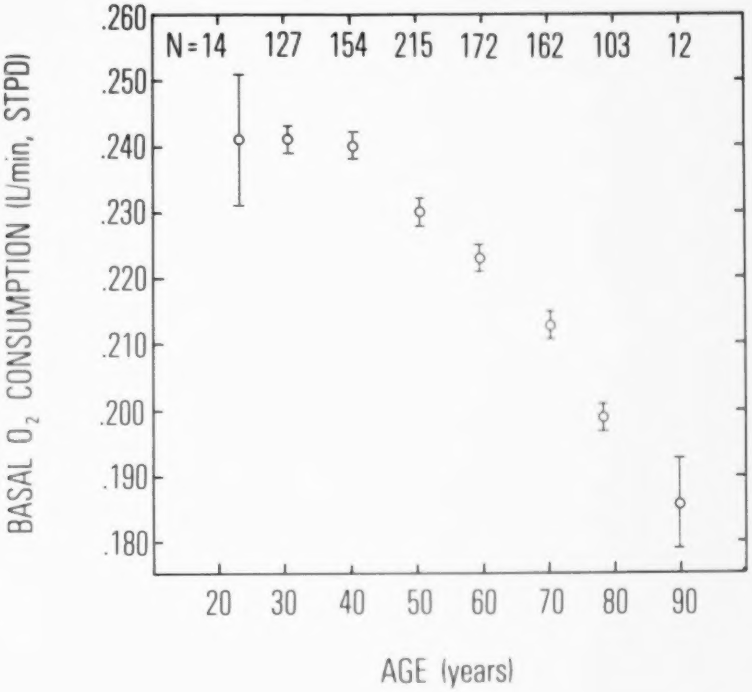


Figure V.1. Basal oxygen consumption (means ± SEM) for men grouped by age. From Tzankoff and Norris (1977).

muscle mass was lower for each older age-group. When basal oxygen consumption attributable to muscle was subtracted from whole-body basal oxygen consumption for each subject, the remainder showed no age-related decrease (Fig. 3). A linear relation was also found between anthropometrically derived lean body mass and 24-hour creatinine excretion for men up to age 65; one g of creatinine excreted in 24 hours was attributed to each 24.8 kg of muscle mass (Fig. 4). It was concluded that diminishing muscle mass may be wholly responsible for the age-related decrease in BMR. These results confirmed an earlier study which had indicated that the fall in basal oxygen consumption with advancing age was primarily a reflection of a loss of functioning tissue estimated from analysis of body water compartments (Shock et al., 1963).

Blood-lactate levels after exercise. Production of energy for muscular work occurs by two processes: aerobic oxidation and anaerobic glycolysis. Anaerobic glycolysis, which produces the smaller proportion of the body's energy, is accompanied by the production of lactate. During exercise, lactate diffuses rapidly from the muscles and is distributed throughout the body by the circulatory system. Most investigators agree that the maximal blood levels of lactate produced by vigorous exercise are generally higher in individuals who are physically fit, but there is disagreement about the best time after exercise to measure lactate and the optimal way to obtain blood samples for its measurement. Most important, there is no information on whether older individuals reach maximal lactate values within the same time as younger ones. To answer this

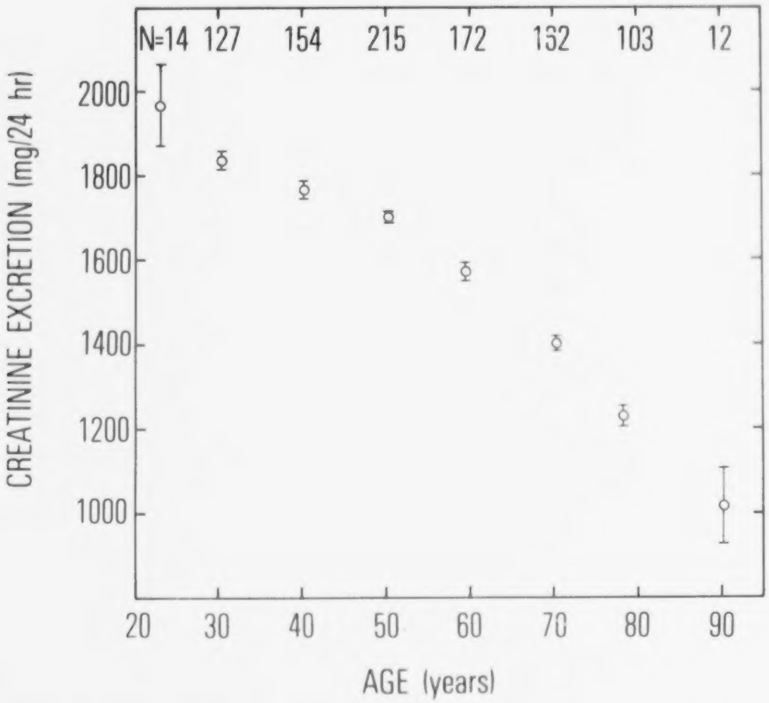


Figure V.2. Values of 24-hr creatinine excretion (means ± SEM) for men grouped by age. From Tzankoff and Norris (1977).

question, a study was conducted to determine whether lactate-distribution kinetics is influenced by age (Tzankoff and Norris, 1979).

The subjects of the study, consisting of 180 men selected from BLSA participants, were grouped in six age decades from the 20s to the 70s. None of the volunteers had clinically detectable cardiovascular disease. Each was given a multistage treadmill test designed to include measurements of maximal aerobic capacity. After an initial two-to-three-minute warmup, the treadmill grade was raised by 3% increments every two minutes until the subjects reported exhaustion. Blood for lactate analysis was drawn from an indwelling catheter at 3, 5, and 7 minutes after termination of the exercise only from subjects who, in the judgment of experienced observers, had exercised maximally.

Blood-lactate levels after maximal exercise were progressively lower with age. The fact that mean values for the youngest men showed no significant differences with sampling time suggests that the diffusion of lactate from muscle and its distribution through the body were complete by the third minute of recovery. In men 30+ years

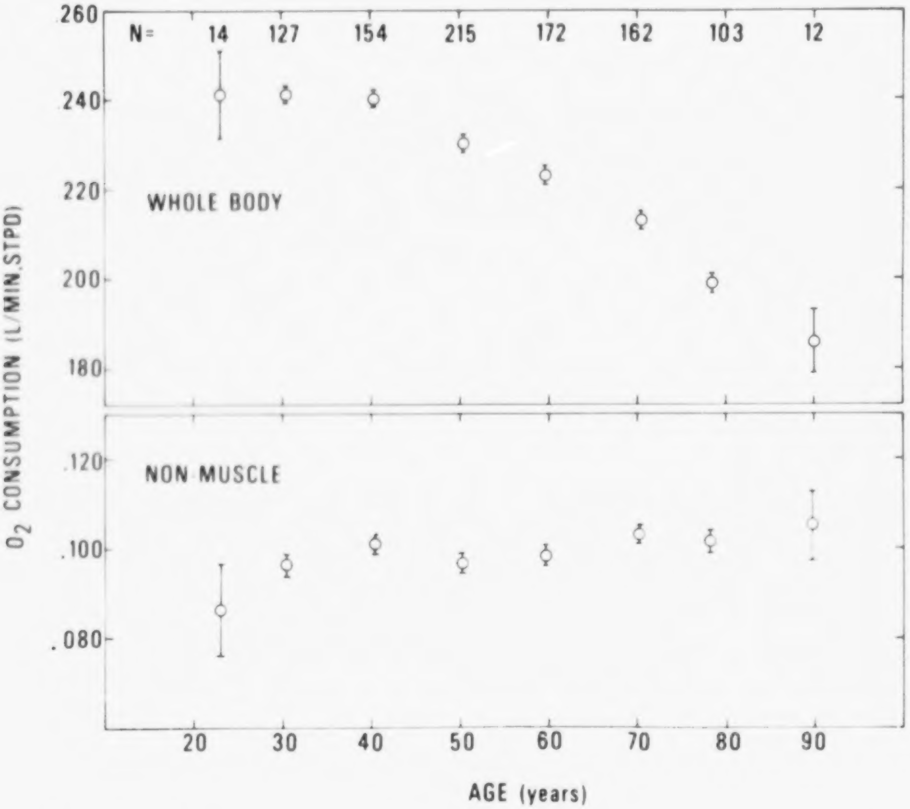


Figure V.3. Lower panel: means ± SEM of calculated nonmuscle oxygen consumption (see text) for subjects grouped by age. Upper panel: same data as in Fig. 1 shown here for reference. From Tzankoff and Norris (1977).

old, mean blood-lactate levels at three minutes of recovery were lower than at five minutes. Although lactate concentrations tended to plateau by the fifth minute in individuals in their 30s and 40s, they continued to rise through the seventh minute in men in their 50s and 60s.

The data suggest a progressive age-related decrease in the rate of diffusion of lactate from muscle and/or its distribution throughout the body, which may contribute to the prolonged recovery seen in old subjects. Although maximal lactate concentrations are best quantified by serial sampling during recovery, this approach is too expensive for most studies. The alternative, for men under 70, is to measure blood lactate in a single sample drawn at the time during recovery from exercise most appropriate for a given age group. Thus blood should be obtained after five minutes of recovery in men up to age 50, and at seven minutes in those between 50 and 70. Variability among men over 70 years of age precludes the use of single samples.

Body dimensions and fat. A radiographic study was undertaken to demonstrate the role of subcutaneous fat in altering external body dimensions in adult males *(Borkan and Norris, 1977).* The sample consisted of 699 men aged 20 to 92 years who were studied between 1958 and 1973. Each 7-x-17-inch soft-tissue radiograph contained views of seven body sites in the trunk and limbs. Measurements were made of skin and fat combined because radiographic differentiation of the two is difficult. Sites of fat measurements on the trunk were bony landmarks such as the top of the greater trochanter; for the calf and forearm, the widest part of the limb was used. When not precluded by factors such as improper body positioning or film handling, fat

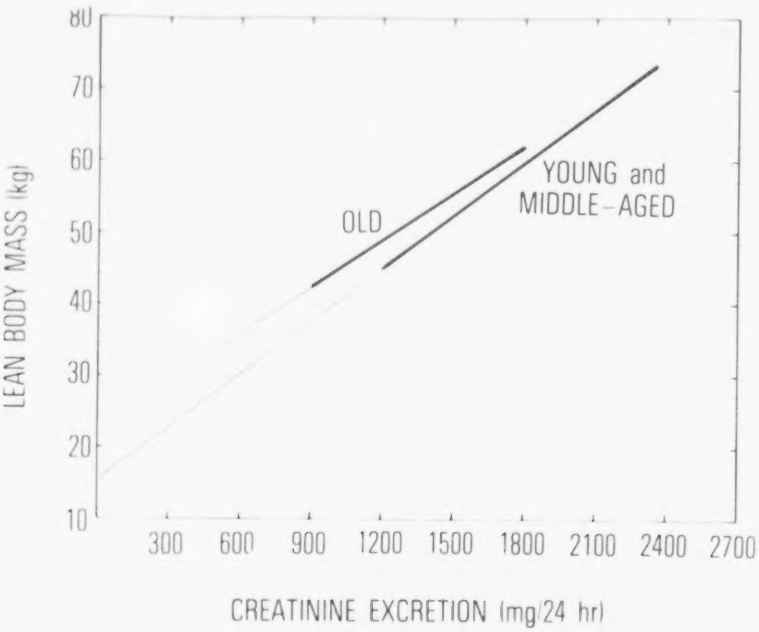


Figure V.4. Relation between lean body mass and creatinine excretion for 259 older (> 65 yr) and 670 younger (up to 64 yr) adult men.
From Tazankoff and Norris (1977).

measurements were made at the following locations: anterior calf, posterior calf, medial calf, lateral calf, lateral to greater trochanter, lateral to top of greater trochanter, lateral to anterior-superior spine of iliac crest, lateral to top of iliac crest, lower part of thorax (lowest rib), medial arm, and lateral arm. A number of new variables were also calculated from other data, including anthropometric body circumferences at sites corresponding to each radiographic fat measurement, height, weight, age, and biochemical assessments of extracellular water and total body water. The new variables included body diameters, muscle plus bone areas for the arm and leg, percentage of body water, percentage of fat, fat-free percentage of the body, and fat-free weight.

The weight of total body fat in this cross-sectional sample was relatively constant with age, but the average fat-free weight was lower in older subjects. In the trunk, data

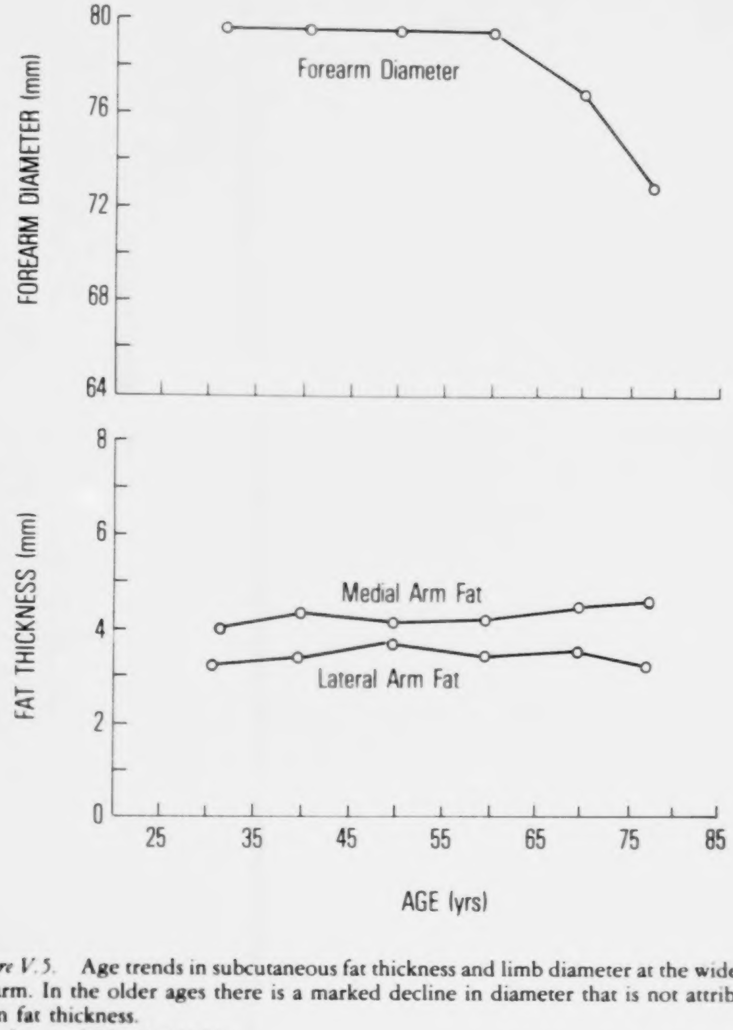


Figure V.5. Age trends in subcutaneous fat thickness and limb diameter at the widest part of the forearm. In the older ages there is a marked decline in diameter that is not attributable to change in fat thickness.
From Borkan and Norris (1977).

from the soft-tissue radiographs revealed that, on the average, subcutaneous fat increases in the region of the greater trochanter but decreases in the abdominal region through middle age. Abdominal diameter increases during this period, however, perhaps as a result of enlargement or sagging of the abdominal contents. The decline in the diameter of calf and arm while fat remains relatively stable suggests loss of lean tissue with age (Figs. 5,6). The findings in this study generally agreed with earlier findings that age changes in body dimensions that result in thin extremities and thicker trunk are only partly attributable to fat redistribution. A part of the age change is the result of tissue loss.

Alterations in bone. Bone loss with increasing age is a universal phenomenon in adults of both sexes, although the rate of loss and the total bone mass lost are greater in

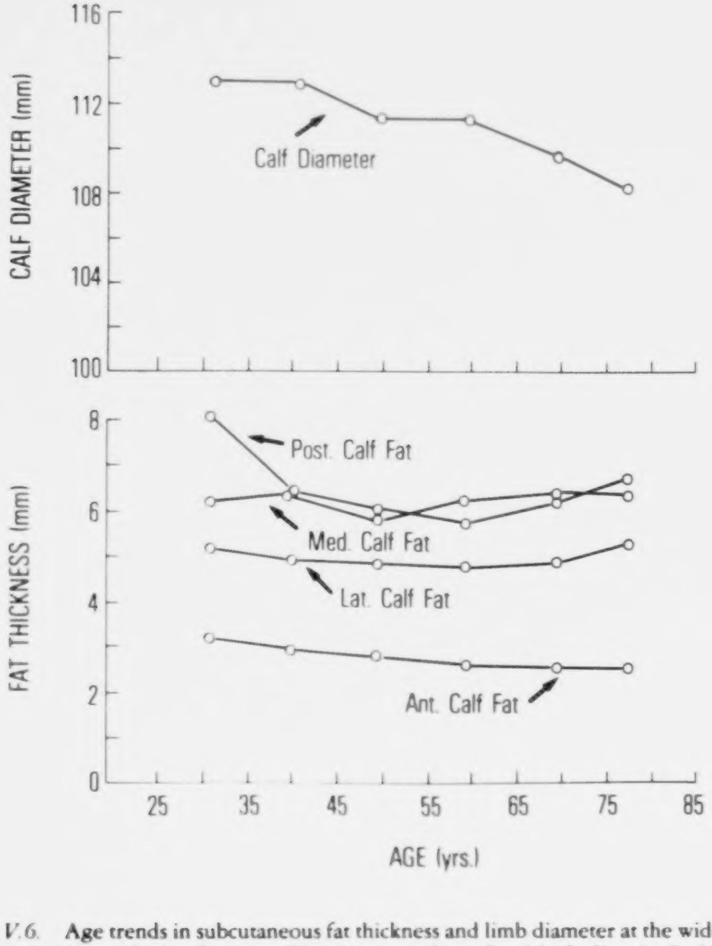


Figure V.6. Age trends in subcutaneous fat thickness and limb diameter at the widest part of the calf. Similarities with the forearm (Fig. 5) are indicated, with decline in limb diameter unaccompanied by loss of fat.
From Borkan and Norris (1977).

females than in males. A study (Plato and Norris, 1980) was undertaken to a) compare, cross-sectionally, the mean values of various bone measurements of the second metacarpal in subjects of different ages; b) investigate possible bilateral differences in bone measurements; and c) determine possible associations between bone measurements and grip strength. Radiographs of the right and left hands were obtained for 236 male BLSA participants whose age range was 25 to 95 years ($\bar{x}$ age = 58.6 yr). The total width, medullary width at the midshaft, and the length of the metacarpal bones were measured from the radiographs, and the data were distributed in seven decade groups between 20 and 89 years of age as well as a 90+ group. Grip strength was measured with a Smedley hand dynamometer.

Age-group comparisons showed that total width and length of the second metacarpal do not change significantly with age after skeletal maturity, while medullary width and area increase and combined cortical thickness decreases significantly. In all age groups the right second metacarpals were wider and longer, and had a thicker cortex, than the left. Medullary width is positively correlated with age, while bone size, total width, and length show no such correlation. After adjustment for age and weight, age and height, and age and surface area, grip strength showed a positive correlation with total width, cortical thickness, and cortical area, but not with medullary width (which is an indicator of bone loss) or length.

In a related study (Plato et al., 1980), 235 normal male BLSA participants were classified as right-handed, left-handed, or ambidextrous on the basis of grip-strength performance. Their left and right hands were also radiographed, and the measurements of the second metacarpal bones were interpreted in terms of hand dominance. Results indicated that the right-hand measurements are generally higher than those of the left hand, regardless of hand dominance, and suggested an inherent tendency to greater bone mass in the right hand. It was postulated that differential stress placed on the bone by the muscles of the right hand in right-handed persons enhances this inherent difference in favor of the right hand to a degree that is statistically significant. In the left-handed, the bilateral difference diminishes because of the increasing stress on the left-hand bone. These findings were confirmed in a follow-up study (Plato and Purifoy, 1982) with an expanded sample of 448 bilateral x-rays from male and 176 from female participants, although in the follow-up study lateral functional hand-dominance criteria, rather than grip strength, were used to determine handedness.

Osteoarthritis of the hand is much more prevalent and more severe in its manifestation in the distal than in the proximal interphalangeal joints. The present system reports osteoarthritis of the hand in most severely affected joints regardless of their position, with the result that the expression of the disease in the proximal interphalangeal and metacarpophalangeal joints is often ignored and underreported. To remedy this methodological insufficiency, a new system of recording the presence and severity of osteoarthritis in all three joints was applied to a study of its age-specific prevalence in 903 BLSA participants, including 228 under 40 years of age, 376 aged 40 to 59 years, and 299 aged 60+ (Plato and Norris, 1979a). The left hands were radiographed to include the digits, the wrists, and the distal heads of the ulna and radius, and the x-ray films were evaluated and graded. Osteoarthritis was more prevalent in older age groups regardless of the joint or digit examined, was more prevalent in the distal than in the proximal joints, and occurred most frequently in the little finger and least frequently in the thumb. It was more severe and its onset earlier in the distal than in the proximal interphalangeal or metacarpophalangeal joints. An unforeseen but statistically verified preponderance of digits with osteoarthritis in both their distal and proximal interphalangeal joints suggests either a common etiology or that the presence of osteoarthritis in the distal joint enhances the likelihood of its development in the proximal joint of the same digit.

3. Nutrition

Caloric intake—diet. In a study of the effect of age on nutrient intakes and energy expenditure (McGandy et al., 1966), intakes of calories and various nutrients were estimated from the daily diet records maintained for one-week periods by 252 healthy male BLSA subjects aged 20 to 99. Because of their high educational level and income, socioeconomic influences on nutrient availability were minimal. Estimates of physical activity were made from detailed interviews, and basal oxygen uptake was also measured. The oxygen required daily for the activities reported by each subject was calculated from estimates of oxygen required for the activity as reported in the literature (Tab. 2). The daily energy expenditure (calculated calories for activity + measured calories for BMR) and its relation to total caloric intake (Tab. 3) were

Table V.2. Calories Expended during Various Physical Activities

Activity	Energy Cost[a] (C/kg per min)
Sitting	
"Active" sitting (writing, talking, etc.)	.027
"Quiet" sitting (reading, watching TV, etc.)	.025
Unspecified active or quiet	.026
Driving car	.035
Eating	.027
Standing	
Unspecified	.03
At drawing board	.035-.04
Teaching, lecturing, etc.	.045-.05
Lying	
Awake	.022
Asleep	.02
Dressing	
Washing, shaving, etc.	.04 (.028-.05)
Walking	
Slow	.04
Moderate	.05 (.04-.06)
Fast	.06 (.06-.08)
Housework	
Making beds	.06
Preparing food, washing dishes	.04
Sweeping floors	.05
Mopping, scrubbing, etc.	.06-.067
Washing windows	.06
Unspecified or miscellaneous housework	.05
Shopping for groceries	.05
Waiting while wife shops for groceries	.025-.04
Caring for young children (feeding, dressing, etc.)	.06
Stairs	
Ascending	.056/flight
Descending	.032/flight
Both ascending and descending (1 flight = 12-16 stairs)	.088/flight

Table V.2. Calories Expended during Various Physical Activities—(*Cont'd.*)

Activity	Energy Cost[a] (C/kg per min)
Gardening	
Dig with hoe	.04-.07
Weeding	.05-.07
Transplanting	.07-.08
Miscellaneous gardening	.06-.08
Mow lawn with hand mower	.09
Mow lawn with power mower	.07
Mow lawn with self-propelled power mower	.06
Mow lawn with riding mower	.045
Driving tractor without attachments	.038
Driving tractor with attachments	.05
Tending greenhouse plants in home or laboratory	.035
Tending greenhouse—commercial	.06-.075
Miscellaneous farm chores	.06
Household Maintenance	
"Unspecified"	.06
"Heavy"	.07-.09
Hand saw	.09
Power saw, drill, etc.	.04
Lay flooring, measure wood, etc.	.07
Paint walls	.07
Paint objects	.04
Repair gadgets, radio mechanics, etc.	.04
Auto repairs	.05-.06
Chopping wood	.09
Sports and Leisure	
Conversing, entertaining	.027
Slow dancing (waltz, fox-trot, etc.)	.06-.065
Fast, vigorous dancing (twist, polka, etc.)	.07-.08
Playing cards	.026
Playing organ or piano	.045
Playing stringed instrument	.04
Playing brass or woodwind	.045
Marching and playing in band	.05
Golf, with caddy	.06
Golf, carrying bag	.075-.08
Golf, unspecified	.07
Swim, vigorous (race)	.07-.08
Swim, relaxed	.06
Ping pong	.06
Bowling	.035
Fishing	.027-.04
Attending sporting events	.03-.05
Singing	.03
Cycling	.05
Fast cycling (race)	.09
Pitching horseshoes	.045-.06
Play baseball	.07
Calisthenics (unspecified)	.07
Playing with young children	.06
Running	.09
Tennis	.09

---

[a] In each case, .02 cal represents basal energy expenditure. If S's own basal rate is used, subtract .02 from figures above (e.g., lying awake = .022 C/kg per min + basal, sitting = .006 + basal, etc.)

From McGandy et al. (1966)

Table V.3. Total Daily Intakes of Various Nutrients in Men of Different Ages

Age (Yr)	20–34	35–44	45–54	55–64	65–74	75–99
N	13	50	52	50	50	37
Height (cm)	180.0 ±4.97[a]	177.8 ±4.58	177.1 ±3.02	175.7 ±4.10	174.4 ±4.03	172.3 ±3.80
Weight (kg)	74.5 ±4.41	77.8 ±6.94	77.6 ±5.16	77.2 ±6.98	77.7 ±7.41	70.9 ±5.79
Total calories	2688 ±584	2639 ±548	2454 ±432	2332 ±345	2297 ±498	2093 ±441
Protein (g)	105 ±16.6	102 ±19.5	98 ±20.3	92 ±17.7	92 ±21.6	81 ±19.1
Fat (g)	123 ±30.6	123 ±29.7	116 ±24.2	105 ±21.8	99 ±30.3	86 ±24.5
Carbohydrate (g)	279 ±73.4	265 ±82.5	240 ±61.9	237 ±52.7	256 ±68.9	244 ±64.6
Alcohol (g)	12	15	14	17	8	9
Calories from protein (%)	15.9 ±1.7	15.6 ±2.1	16.1 ±2.7	15.8 ±2.4	16.1 ±2.7	15.5 ±2.5
Calories from fat (%)	41.0 ±3.2	42.0 ±4.7	42.5 ±4.6	40.3 ±5.2	38.5 ±6.4	36.9 ±6.5
Calcium (g)	1.29 ±0.83	1.00 ±0.38	0.79 ±0.32	0.74 ±0.26	0.91 ±0.44	0.89 ±0.42
Iron (mg)	16.2 ±4.3	14.5 ±2.6	15.0 ±2.7	14.3 ±3.3	14.0 ±2.5	12.3 ±2.8
Vit. A (I.U. x 10^{-2})	119 ±70.1	78 ±51.2	78 ±35.0	79 ±56.8	89 ±45.8	81 ±36.6
Thiamine (mg)	1.67 ±0.77	1.38 ±0.33	1.28 ±0.32	1.20 ±0.28	1.35 ±0.37	1.20 ±0.28
Riboflavin (mg)	2.70 ±1.26	2.21 ±0.72	1.91 ±0.56	1.83 ±0.64	1.98 ±0.74	1.87 ±0.59
Niacin (mg)	23.1 ±6.2	20.1 ±4.1	20.0 ±4.7	18.5 ±4.4	18.0 ±4.6	15.0 ±4.5
Ascorbic acid (mg)	106 ±55.4	107 ±44.6	106 ±44.2	115 ±55.7	142 ±62.0	119 ±51.3
Calories from saturated fat (%)	16.4 ±1.1	16.7 ±2.5	16.7 ±2.3	15.5 ±2.8	14.8 ±2.4	14.4 ±3.0
Calories from polyunsaturated fat (%)	5.4 ±1.3	5.2 ±1.1	5.4 ±1.0	5.3 ±1.2	5.2 ±2.5	4.8 ±1.3
Cholesterol (mg)	580 ±281	610 ±200	620 ±190	600 ±214	540 ±220	480 ±181

[a] Mean ± S.D.

From McGandy et al. (1966)

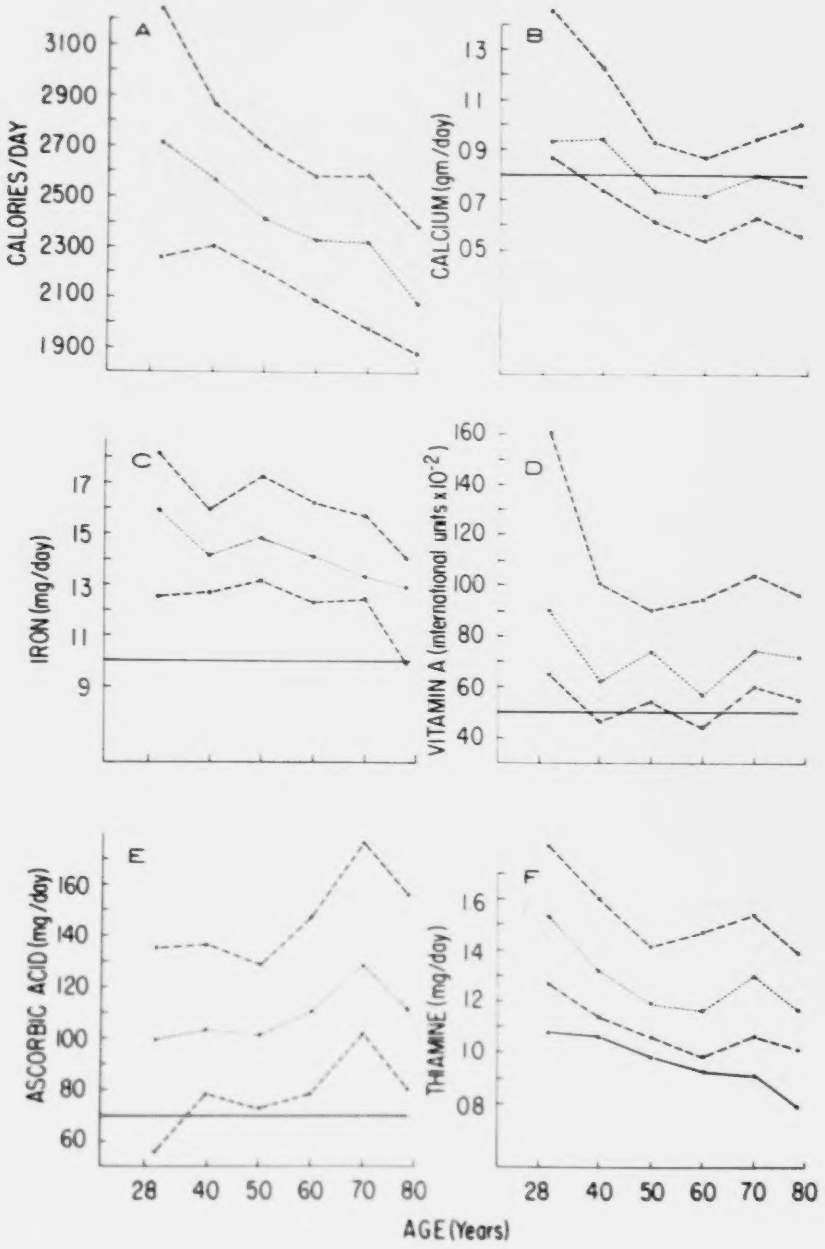


Figure V.7. Total daily intakes of calories (A), calcium (B), iron (C), vitamin A (D), ascorbic acid (E), and thiamine (F) in men of different ages. The medians are represented by the dotted lines and the first and third quartiles by the dashed lines. Solid lines represent National Research Council recommended allowances.
From McGandy et al. (1966).

measured in subjects of different ages. Older subjects were found to consume markedly fewer calories than younger subjects (Fig. 7A). Figure 7 shows that the pattern of average age differences varied considerably among nutrients. None followed the almost linear drop observed in the average decrement in total calories consumed (Fig. 7A). In the case of vitamin A the decrement in intake was confined to the age span of 20 to 40 (Fig. 7D); for calcium, the span was from 20 to 50 (Fig. 7B). In contrast, the average intake of ascorbic acid increased after age 50. Except perhaps for calcium, the daily allowances recommended by the National Research Council were met by most of the subjects.

Although the percentage of calories from protein was remarkably constant with age (Fig. 8A), calories derived from fats dropped from 42% in the 45-54-year age group to 36% in the 80-year-old group, while the contribution of calories from

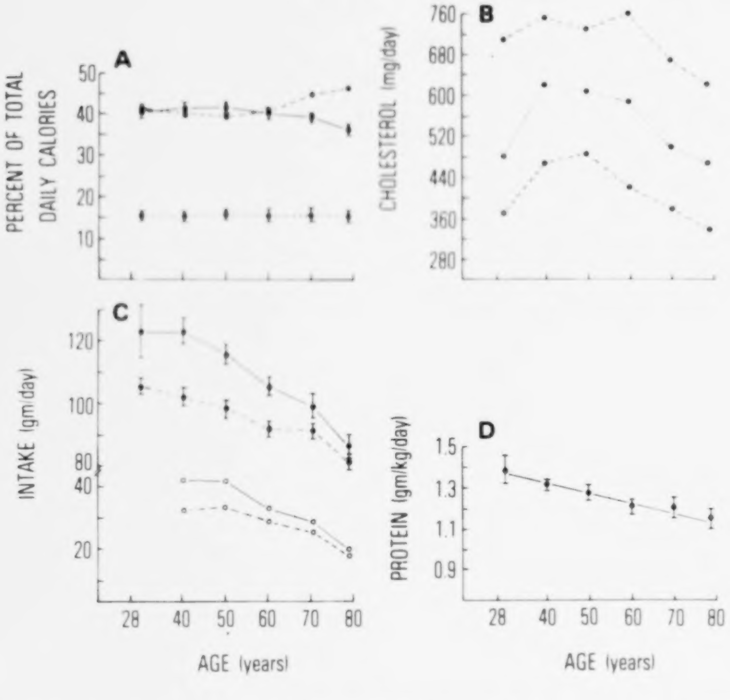


Figure V.8. A. The percentage of the total daily caloric intake derived from fat (——), carbohydrate (----), and protein (...) in men of different ages. Vertical bars represent SEM. Correlation coefficients for regressions on age of percentage of calories from fat and from protein were –.37 ($p < 0.01$) and –.004, respectively.

B. Total daily intakes of cholesterol in men of different ages. The medians are represented by the dotted line and the first and third quartiles by the dashed line.

C. Total daily intakes of fat (●——●) and protein (●----●) in men of different ages. Vertical bars represent SEM. The lower curves represent the intakes of fat (○——○) and protein (○----○) derived from meats exclusive of poultry and fish.

D. Total daily intakes of protein per unit of body weight in men of different ages. Vertical bars represent SEM. Correlation coefficient = –.242 ($p < 0.01$).
From McGandy et al. (1966).

carbohydrates increased correspondingly. A decrement in cholesterol intake was related to and largely accounted for by the reduced use of meats (except poultry and fish) in the older subjects (Fig. 8B).

Basal metabolism decreased by 5.23 cal/day per year of age (Fig. 9B), while total caloric intake fell 12.4 cal/day per year (Fig. 9A). The difference, 7.17 cal/day per year, must be related to a reduction in calories required for other purposes, including physical activity. When expressed in relation to body weight, this decrease in non-basal energy expenditure showed a plateau from age 60 (Fig. 9D). There was close agreement between the total caloric requirement per day (activity cal + basal cal) and the total caloric intake calculated from dietary diaries.

Vitamin B_6. The vitamin-B_6 status of 617 men ranging in age from 18 to 90 years was investigated by use of plasma pyridoxal phosphate (PLP) and glutamic-oxaloacetic transaminase (GOT) concentrations to assess B_6 nutrition and to determine the effects of vitamin-B_6 supplements (Rose et al., 1976) (Fig. 10). The studies provide the most extensive normative data to date on the vitamin-B_6 status of normal men in the adult years. Almost half the men in the 60s and 70s, and one fourth of the younger men and

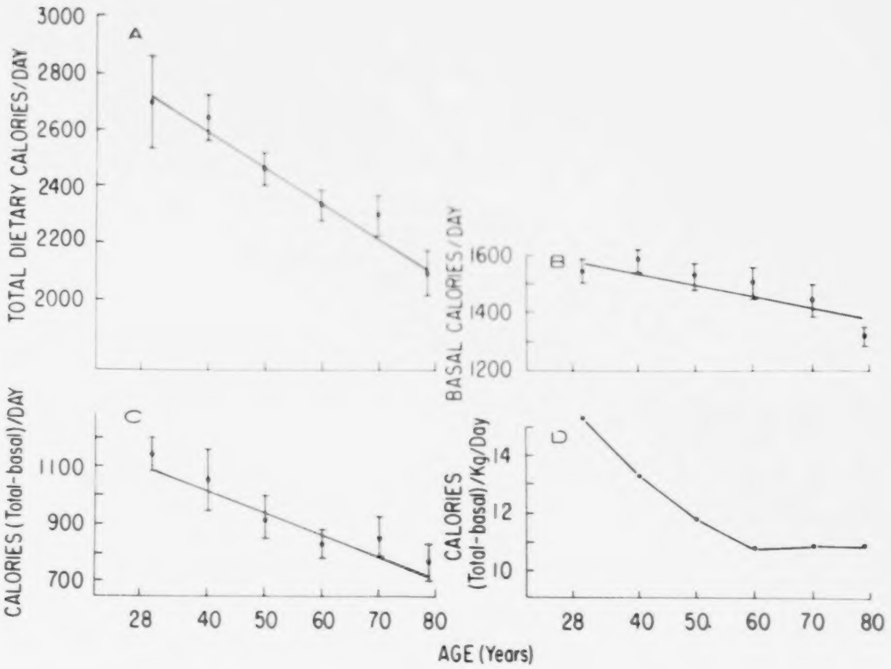


Figure V.9. Mean total daily caloric intakes (A), basal metabolic rates (cal/24 hr) (B), energy expenditures (C), and energy expenditures per unit of body weight (D) in men of different ages. Vertical bars represent standard errors of the means. Correlation coefficients for regressions on age of total calories, basal calories, and total-basal calories were −.374, −.374, and −.231, respectively; all were statistically significant ($p < 0.01$).
From McGandy et al. 1966).

of those over 80, were taking a vitamin supplement containing B_6. The average PLP level for the 203 men taking vitamin supplements was 20.5 ± 1.0 ng/ml, compared to 12.3 ± 0.3 ng/ml in the 414 not taking supplements. Subjects not taking B_6 supplements showed a statistically significant decrease in plasma PLP with age (0.9 ng/ml per decade of age). Although subjects taking supplements also showed a decline in PLP with age, the decline was not statistically significant. Plasma GOT levels varied little with increasing age, although they were significantly higher in the group taking vitamin supplements.

Serum albumin and aging. One cellular theory assumes that aging is a consequence of the reaction between metabolically produced free radicals and easily oxidized chemical groups, such as sulfhydryl groups (D. Harman, 1981). Consistent with this theory is the finding that the concentration of sulfhydryl groups in serum declines with age. Most serum-sulfhydryl groups, however, are associated with albumin, and since albumin levels also decline with age, lower serum-sulfhydryl levels in older people may be due either to free-radical oxidation or to lower levels of albumin production. A study was therefore undertaken to determine to what degree age-dependent decreases in serum-sulfhydryl concentration are due to decreased levels of serum albumin (Leto et al., 1970).

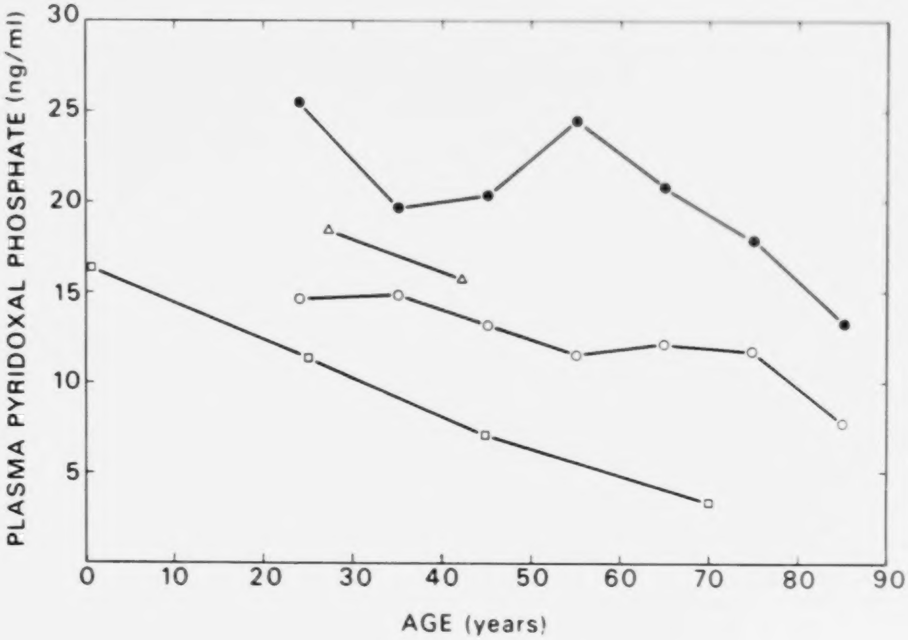


Figure V.10. Plasma pyridoxal phosphate levels reported by Rose et al. for the BLSA and by other investigators. (●) Rose et al., vitamin-B_6 supplemented; (△) Chabner et al.; (○) Rose et al., unsupplemented; (□) Hamfelt.
From Rose et al. (1976).

Fasting blood samples were obtained from 194 apparently healthy male volunteers participating in the GRC's longitudinal program. After assay for total albumin concentration, sera from the samples were fractionated to isolate albumin, which was then assayed for sulfhydryl content.

In men, the concentration of albumin, albumin sulfhydryl, and serum sulfhydryl was significantly lower in the ninth than in the third decade. Measurement of the readily reducible sulfhydryl/albumin ratio in a subsample of the men failed to demonstrate age differences. The data suggest that decrements in the sulfhydryl/albumin ratio result from changes in the redox potential of sera rather than from qualitative age changes in the structure of newly synthesized albumin. The observations failed to provide support either for the free-radical theory of cellular aging (D. Harman, 1981) or, incidentally, for the "error" theory (Medvedev, 1964; Rockstein, 1974) of aging.

4. Drug Metabolism

The effect of aging on drug metabolism was examined in two studies.

Antipyrine. The influence of age, alcohol consumption, caffeine consumption, and smoking on antipyrine metabolism was studied in 307 healthy male subjects aged 18 to 92 years (Vestal et al., 1975). The plasma half-life of intravenously administered antipyrine was 16.5% longer and metabolic clearance rate 18.5% lower in the older group (60–92 yr) than in the younger group (18–39 yr). There was a small, but statistically significant, negative correlation (–0.25, $p < .001$) between age and metabolic clearance rate of antipyrine. Consumption of caffeine and alcohol, as well as use of cigarettes, declined across the age span. Caffeine consumption was positively correlated with the rate of antipyrine metabolism in all age groups. Cigarette consumption was positively correlated with metabolic clearance rate of antipyrine only in the young and middle-aged groups; clearance rates in the older smokers were comparable to those of nonsmokers. No significant relation between alcohol consumption and antipyrine metabolism was found. Multiple regression analysis showed that smoking explained 12% and age only 3% of the variance in the rate at which antipyrine was removed from the blood. The results suggested that habits that differ with age—such as the decline found here in consumption of caffeine, cigarettes, and alcohol—must be taken into account in studies attempting to quantify the effects of aging.

Ethanol. The effect of aging on the distribution and elimination of ethanol was studied in 50 healthy men aged 21 to 81 years who had abstained from alcohol for three weeks (Fig. 11) (Vestal et al., 1977). Ethanol doses of 375 mg/m^2 body-surface area per minute were administered to the subjects by continuous one-hour intravenous infusion. Over the next four hours, blood samples were obtained at intervals of 15 to 30 minutes for measurement of ethanol concentration. The minimal model that satisfied the ethanol distribution and metabolism data was a two-compartment model in which an initial compartment represents blood and interstitial fluid spaces that exchange rapidly with blood, while a secondary and larger compartment represents more slowly equilibrating interstitial fluid and intracellular fluid spaces. Rates of ethanol elimination were not affected by age, but a significant correlation was found between age and the peak blood-ethanol concentration at the end of the infusion period (Fig. 12). Since all subjects received equivalent ethanol doses on the basis of

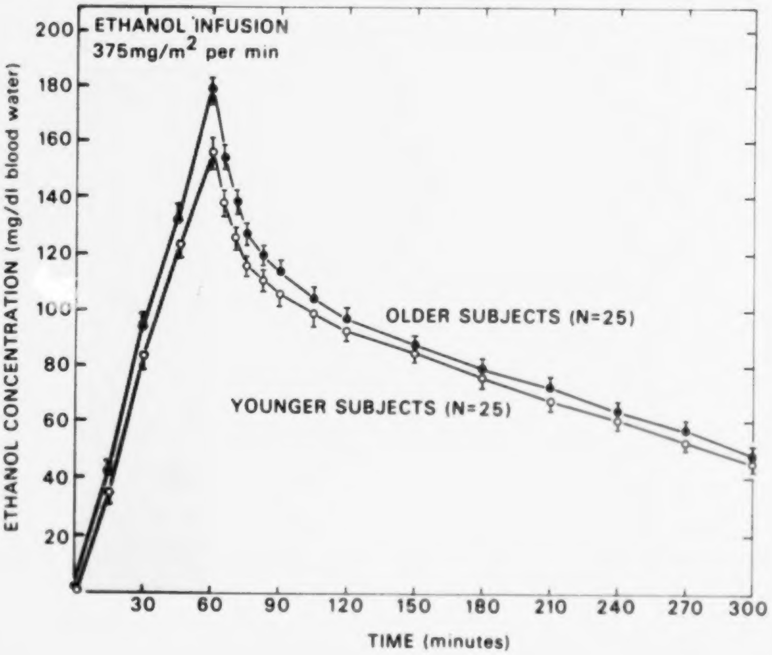


Figure V.11. Mean ethanol concentration in blood water (± SEM) for younger (aged 21–56) and older subjects (aged 57–81).
From Vestal et al. (1977).

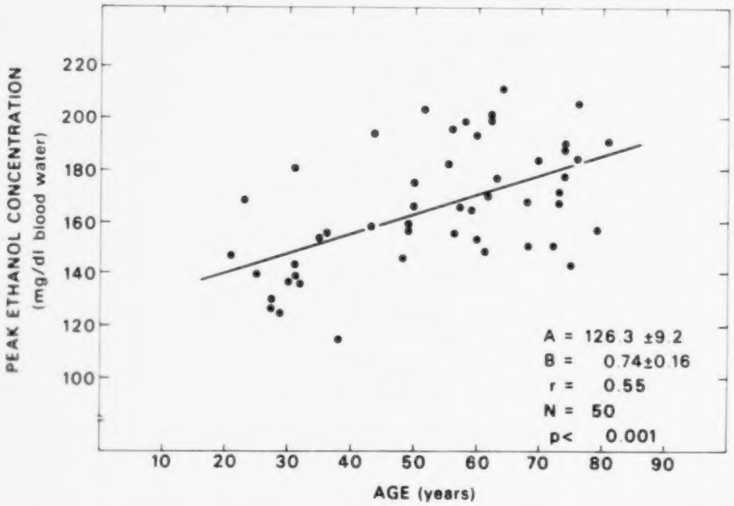


Figure V.12. Correlation with age of peak ethanol concentration in blood water at the end of ethanol infusion. A indicates the intercept and B the slope.
From Vestal et al. (1977).

body-surface area, it was concluded that the higher peak ethanol levels in the body water of old subjects were probably due to smaller volume of body water and lower lean body mass. Age-related changes in body composition are thus important factors in the metabolic and pharmacologic effects of drugs.

5. Pulmonary System

Age differences in pulmonary function. Earlier studies of age differences in lung volumes and maximal breathing capacity were usually carried out in hospital environments, often in subjects whose pulmonary status may not have represented that of active, successful people living at home. In addition, older hospital patients selected for pulmonary studies were seldom drawn from higher socioeconomic groups. A study

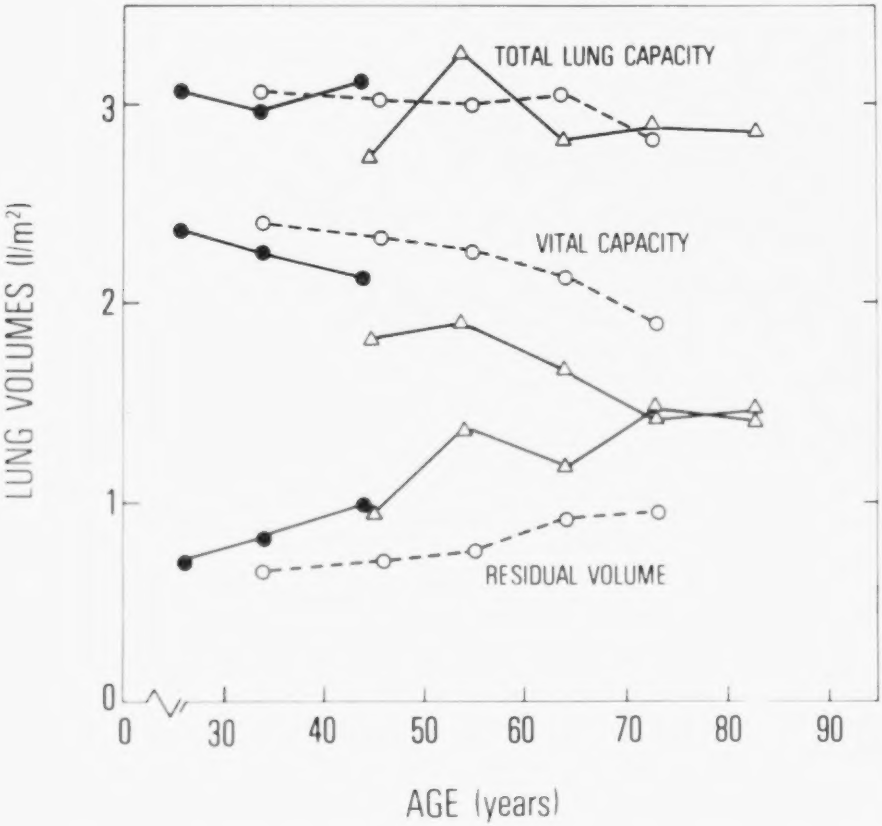


Figure V.13. Lung volumes divided by body-surface area (l/m^2) are plotted against age (yr). A hospital indigent group (△), a hospital staff group (●), and a community-residing group of BLSA participants (○). Average values by age decades (20–29 yr, etc.) are shown.
From Norris et al. (1962).

was undertaken to compare lung volumes and maximal breathing capacity in residents of the Infirmary Division (old people's home), employees, and short-term patients of Baltimore City Hospitals (BCH) (group A) and a sample of the BLSA study population (group B) (Norris et al., 1962).

Group A comprised 135 subjects whose age range was 20 to 89 years. Group B consisted of 166 BLSA participants whose age range was 30 to 79 years. Individuals from group B were generally taller, heavier, and better educated than individuals selected from the BCH population.

In group B, a statistically significant regression on age was found for maximal breathing capacity, vital capacity (both are lower with age), and residual volume (which is higher with age), but not for total lung capacity (Fig. 13). Age thus had a greater effect on a dynamic measurement of lung function that required a coordinated neuro-muscular response—breathing at a maximum rate and volume—than on a measurement of a single response, such as the single maximum expiration required for the measurement of vital capacity. Vital capacity was higher in group B than in group A, whereas residual volume was lower in group B than in group A for the sixth and eighth age decades. Total lung capacity did not differ in the two groups. Residual volume was significantly higher in 40-49-year old hospital-staff members of group A than in members of the same age group in group B. The group differences suggested a relation between pulmonary functional capacity and socioeconomic status (Fig. 14).

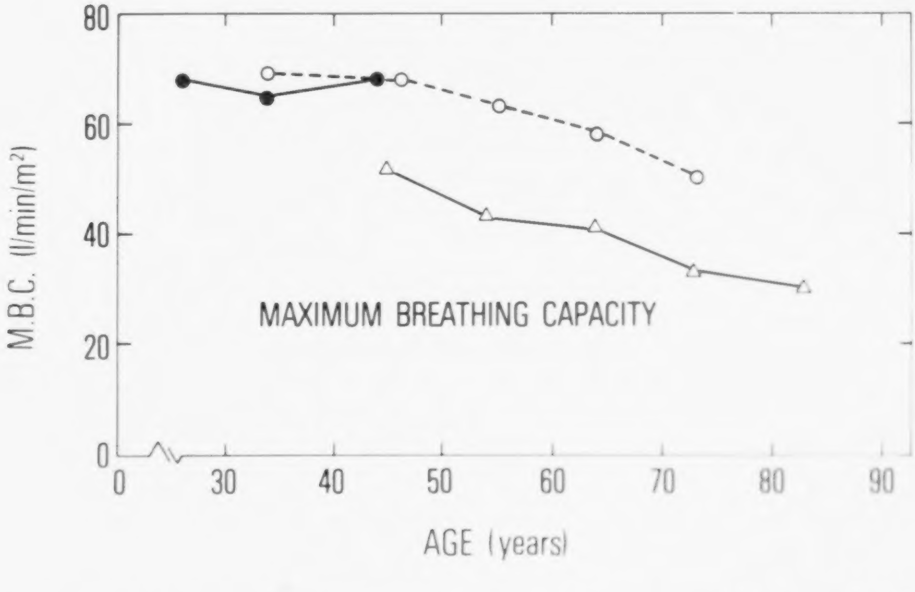


Figure V.14. Maximal breathing capacity in a hospital indigent group (△), a hospital staff group (●), and a community-residing group (○). Average values by age decade (20–29 yr, etc.) are shown.
From Norris et al. (1962).

Table V.4. Mean Values, Range of Values, and Standard Deviations of the Distributions of Lung-Clearance Index by Age Groups

Age in Yr	Mean Value	Range of Values	Standard Deviation	Number of Subjects
29–39	9.17	7.91–11.04	1.02	7
40–49	13.66	10.82–19.80	3.20	7
50–59	12.06	9.62–14.80	2.02	11
60–69	12.87	9.96–18.00	2.20	12
70–90	14.86	11.28–17.36	2.08	10

From *Norris et al. (1964)*

A comparison was made between BLSA participants and residents of the BCH old people's home to re-examine a previous finding that the ratio of ventilation to consumption of oxygen increases with age *(Norris et al., 1964)*. Minute volume and breathing rate were similar for young and old persons in both groups, although vital capacity and maximum breathing capacity declined with age. Minute volume and alveolar ventilation in resting subjects were significantly higher in residents of the old people's home than in BLSA subjects. Nitrogen-washout curves analyzed in 47 BLSA subjects revealed significant age difference in the lung-clearance index (Tab. 4). This measurement indicates that older subjects are unable to maintain uniform ventilation of their lungs. A sample of 108 BLSA subjects who were tested at 18 and 36 months after their first visits showed no longitudinal changes in pulmonary function except an increase in maximum breathing capacity between the first and second visits in all age groups; the increase was attributed to the effects of practice in performing the test.

Age-related differences in the size of the lung compartments were examined in 42 normal males aged 24 to 78 years (Mittman et al., 1965). Measurements of chest-wall and pulmonary compliance used the static and the positive-pressure breathing methods of Heaf and Prime (1955). Chest-wall and total compliance were found to be significantly lower with age, but there was no significant relation between age and pulmonary compliance. A decrease in pulmonary compliance with higher lung volume was greater in young than in old subjects, possibly because of a loss of elastic recoil in the lungs of the elderly or because of age differences in chest-wall compliance.

The uniformity of distribution of pulmonary ventilation was assessed in 117 BLSA participants whose age range was 20 to 103 years (Edelman et al., 1968). The data from nitrogen-washout studies demonstrated that ventilation was significantly less uniform in old than in young men. Ventilation uniformity improved with higher tidal volume only in the older group. Studies of smaller groups in which each subject served as his own control confirmed that old men could improve their ventilation uniformity by breathing deeply. A single forced expiration before nitrogen washout impaired the ventilation uniformity of old but not of young men. The findings suggest that the lungs of old men are more susceptible than those of young men to localized alveolar collapse.

Effects of cigarette smoking on pulmonary function. Significant differences were found in the performances of cigarette smokers and nonsmokers on several spirometric tests (Edelman et al., 1966). The subjects, including 360 men participating in the BLSA and 50 medical and laboratory personnel, ranged in age from 20 to 103 years ($\bar{x}$ age = 52). Subjects were accepted if they had no history or signs of bronchopulmonary disease, cardiovascular insufficiency, or muscular weakness. Smoking habits were assessed by a

Table V.5. The Prevalence of Coronary Artery Disease in the Elderly as Assessed by Resting[a] and Stress Criteria[b]

	Age			
	51–60	61–70	71–80	81–90
N	70	73	36	10
% CAD by resting	13	15	22	20
% CAD by resting *and* stress	24	37	56	50

[a] History of angina or myocardial infarction; abnormal resting ECG, i.e., Minnesota codes 1:1, 1:2, 1:3, or 4:1.

[b] ECG positive for ischemia during maximum exercise treadmill test, i.e., Minnesota code 11:1 or an abnormal thallium scan during maximum exercise but not at rest.

From Gerstenblith et al. (1980)

questionnaire; spirometric data were analyzed by dividing the subjects into four groups based on their smoking histories: nonsmokers, current cigarette smokers, former cigarette smokers, and current and former pipe and cigar smokers. Current cigarette smokers had significantly lower values for vital capacity, maximal ventilation, and maximal expiratory flow rates than a comparable group of nonsmokers. Former cigarette smokers had significantly lower vital capacity than nonsmokers. Pipe and cigar smokers performed as well as nonsmokers in all tests.

Although comparison of former cigarette smokers with nonsmokers and current cigarette smokers did not provide clear evidence for or against reversibility of ventilatory impairment, analysis of the former smokers showed that those with the longest smoking-free interval performed significantly better on several tests of pulmonary function. Within the limitations of a retrospective cross-sectional survey, the findings are consistent with reversibility of pulmonary impairment upon cessation of smoking.

6. Cardiovascular System

Cardiac Performance and Aging: Occult Disease

Previous cross-sectional studies have characterized the effects of aging on a number of cardiovascular functions. Findings included a progressive fall in resting cardiac function with advancing age, a rise in peripheral resistance, and a wide range of individual differences in all measurements (Brandfonbrener et al., 1955). None of these earlier studies was able to control for the presence of occult coronary-artery disease (CAD), nor did the studies consider the average daily level of physical activity. The BLSA population provides a unique opportunity to reassess the effect of aging on the cardiovascular system for at least three separate reasons: Stress testing is used to screen for occult CAD; the elderly participants in whom cardiovascular function is tested maintain a level of physical activity required for independent community living; and recent technological advances that have made possible more accurate non-invasive measurements of cardiac function are currently available for application to BLSA participants.

The value of stress testing to detect occult CAD is illustrated in Table 5 (Gerstenblith et al., 1980). A subset of the population in the sixth to ninth decades was evaluated for CAD by both resting and stress criteria, which included exercise ECG and thallium testing described in Chapter IV. When stress criteria are employed in subjects in the sixth through the ninth decades, CAD is found in roughly twice as many cases, so that its rate in the aged approximates that found in unselected necropsy reports (Gerstenblith et al., 1980). It is likely that the high prevalence of occult disease in the older age range has had a major impact on studies attempting to assess the effect of age *per se* on cardiovascular performance. In the BLSA approach, data from subjects with either clinically overt or latent disease are excluded from studies whose purpose is to specify age effects; the result has in many instances been a perspective on the aging heart that is somewhat different from those of earlier studies: Some decreases in cardiovascular performance with age are due to disease, and the impact of age *per se* may not be as great as has been thought.

Cardiac Function at Rest

Systolic time intervals. Systolic time intervals provide a non-invasive assessment of general cardiovascular performance. In a series of 315 BLSA participants aged 20 to 89 years, a slightly longer ejection period was observed with age (Shaw et al., 1973).

One-dimensional echocardiography. The advent of ultrasound as a tool to test cardiac dimension and motion has permitted a more direct non-invasive assessment of cardiac function. A major determinant of cardiac stroke volume is the end-diastolic filling

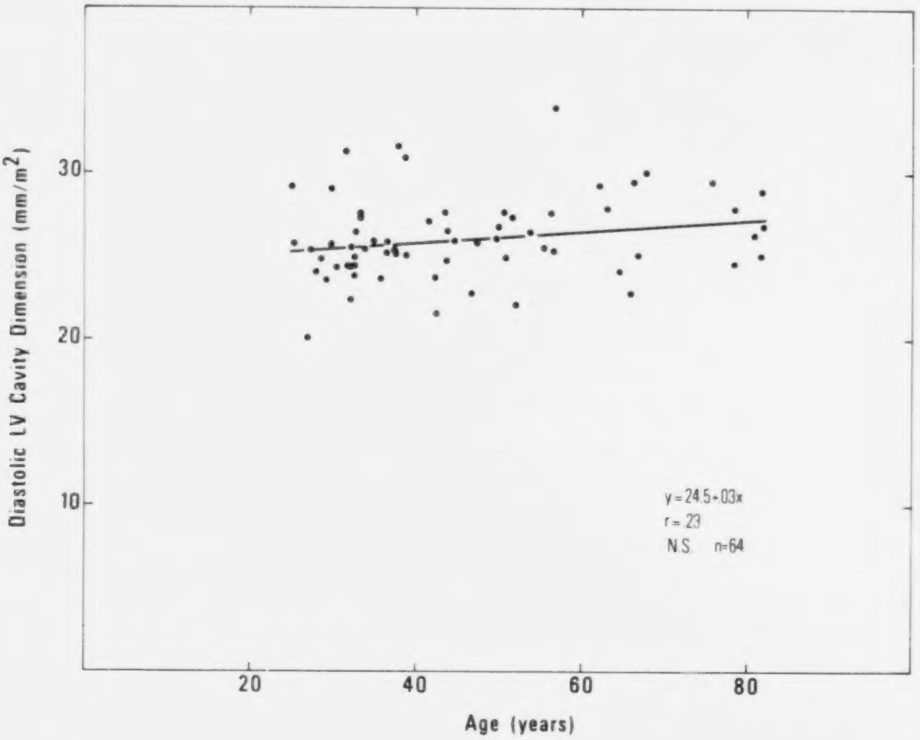


Figure V.15. Linear regression plot depicting the relationship between age and diastolic cavity dimension in healthy participants from the BLSA.
From Gerstenblith et al. (1977).

volume. Studies utilizing M-mode (one-dimensional) echocardiography in BLSA participants have indicated that, at rest, the left-ventricular end-diastolic diameter is not significantly altered with advancing age (Fig. 15) (Gerstenblith et al., 1977), and that ejection fraction index is also not age-related (Fig. 16). In view of the fact that the resting heart rate did not vary with age in these subjects, and if it can be assumed that the heart does not change its shape with age, the results suggest that cardiac output at rest is also not age-related. But since the assumption about shape may not be warranted, one-dimensional echograms permit no firm conclusion about the effect of age on cardiac output.

Two-dimensional echocardiography. Technological advances have resulted in the evolution of an echogram that measures cardiac chambers in two dimensions. Two-dimensional studies in an additional subset (n = 25) of BLSA participants indicated that end-diastolic and end-systolic cardiac areas were not changed with age (VanTosh et al., 1980); since the study found no age-related change in heart rate, it also suggested that resting cardiac output did not vary with age.

Multiple-gated cardiac blood-pool (MUGA) scans. More recently, it has been possible to measure cardiac volumes directly by multiple-gated cardiac blood-pool scans

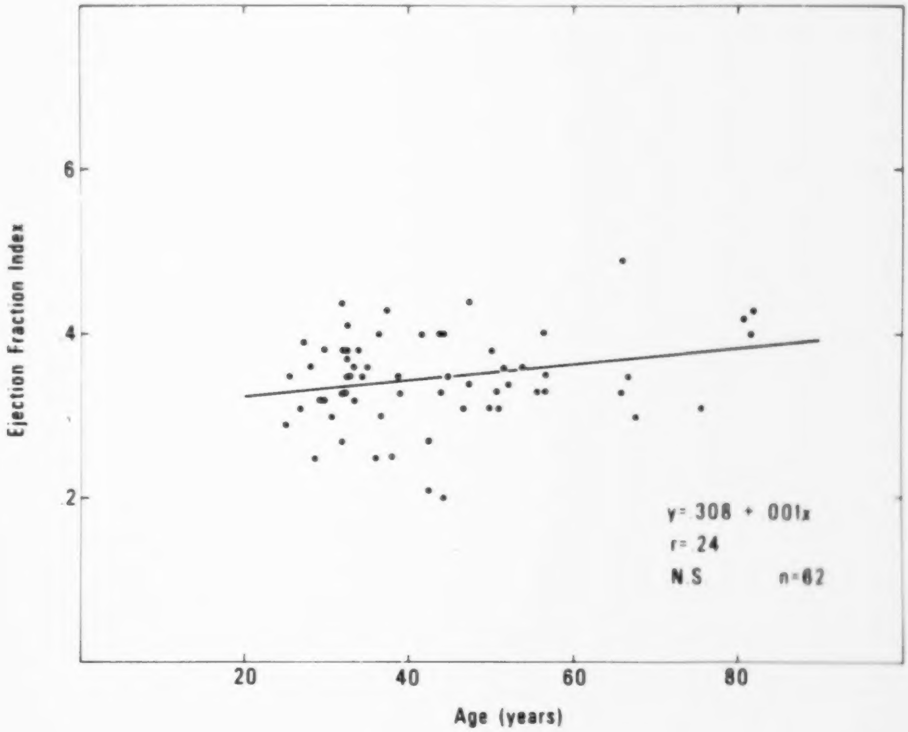


Figure V.16. Linear regression plot depicting the relationship between age and ejection fraction index in healthy BLSA participants.
From Gerstenblith et al. (1977).

Table V.6. Effect of Adult Age on Resting Cardiac Function

	Population A	Population B
	Institutionalized Unscreened for Occult CAD[a] Age range, 19–86 yr	Active in Community Life Screened for Occult CAD[b] Age range, 24–79 yr
Heart rate	Slight decrease	No effect
Stroke volume	Decrease	No effect
Stroke-volume index	Decrease	No effect
Cardiac output	Decrease	No effect
Cardiac index	Decrease	No effect
Peripheral vascular resistance	Increase	No effect
Peak systolic blood pressure	Increase	Increase
Diastolic pressure	No effect	No effect

[a]From Brandfonbrener et al. (1955)
[b]From Rodeheffer et al. (1981)

(MUGA) described in Chapter IV. Analysis of studies in the first 37 participants (Rodeheffer et al., 1981) confirms the interpretation of the earlier studies with ultrasound, that no relation to age is found when end-diastolic, end-systolic, stroke volume, or cardiac output is measured at rest. A comparison of the hemodynamic profile of these participants with that of institutionalized subjects of an earlier non-BLSA study is given in Table 6. With the exception of the age-related increase in systolic blood pressure noted in both studies, the age-related differences in cardiovascular function found in the previous study were not observed in BLSA participants. Although the earlier study was more heavily weighted on both age extremes, the marked differences between the two groups between the ages of 30 and 80 years cannot readily be attributed to a difference in age range or in methodology. The contrast must thus result from differences in the populations studied. While this might in part be a birth-cohort effect of unspecified cause, a more attractive hypothesis would point to the likelihood of cardiovascular "deconditioning" secondary to convalescence or bed-rest, and occult coronary disease in the earlier sample.

Ventricular hypertrophy. Although at rest cardiac output and its hemodynamic determinants—filling volume, ejection fraction, and heart rate—are not age-related, the increase in systolic blood pressure with advancing age necessitates an increase in cardiac work. The increase in systolic pressure apparently results from age-related stiffening of the vasculature (Lakatta, 1979).

In order to develop and maintain greater systolic pressure, ventricular wall stress must increase. Because of the inverse relation between wall stress and fiber shortening, however, the increase in wall stress compromises shortening and thus the ejection of blood. To maintain normal shortening in the face of an increased pressure load, the heart can hypertrophy, i.e., thicken its wall, since wall stress is force per unit thickness. The presence of cardiac hypertrophy with advancing age has been sought in 67 healthy BLSA participants. Gerstenblith et al. (1977) have demonstrated a mild age-related increase in left-ventricular wall thickness in men over the age range of 25 to 80 years (Fig. 17) that may be considered an adaptation to maintain normal systolic function. Although the increase in diastolic wall thickness may be a major factor in producing the progressive age-related limitation of cardiac filling during the early diastolic period (Fig. 18), the fact that end-diastolic volume is not compromised renders it unlikely that the reduced filling rate at rest has physiologic significance.

Cardiovascular Response to Stress.

The fact that overall cardiac performance at rest is not altered by age in normal man by no means indicates that performance during stress is unaffected. Classic studies of exercise physiology have indicated that maximum exercise performance diminishes with age in apparently healthy persons. Age-related declines have been identified in maximum aerobic capacity, heart rate, stroke volume, cardiac output, and arteriovenous O_2 differences (see review by Gerstenblith et al., 1976). These comparisons, however, are made not at standard exercise levels across all ages but at the voluntary maximal exercise level in each subject, which is substantially reduced with age. Since no plateau is evident in the workload curves, it cannot be ascertained whether maximum cardiac performance was achieved in the elderly, or whether the limitation in workload with advanced age might have been due to other factors.

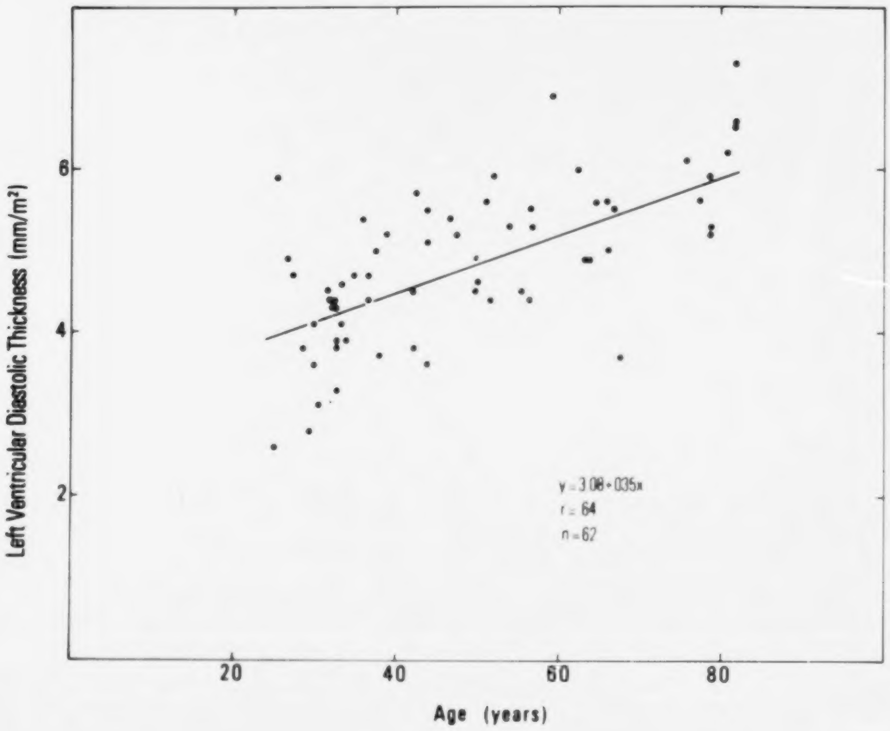


Figure V.17. Linear regression plot depicting the relationship between age and left-ventricular diastolic wall thickness.
From Gerstenblith et al. (1977).

Consideration of the influence of occult disease is even more important during exercise than in studies of the cardiovascular system at rest. In fact, an age-dependent increase in ECG abnormalities during exercise, which may in some instances be indicative of coronary disease, has been observed in some of the classic studies (Montoye, 1975; Granath et al., 1964). Furthermore, the status of physical conditioning is also a major determinant of the exercise response. A true assessment of the effect of aging *per se* on the cardiovascular response to exercise, then, requires strict control of both parameters. The cardiovascular response to exercise has been measured by the MUGA scan in BLSA participants who had both normal ECG and normal thallium tests during treadmill exercise. The results obtained in the first 37 consecutive volunteers for this exercise protocol, like the findings at rest, differ in many ways from previous studies (Gerstenblith et al., 1976) that have examined the effect of age on the cardiovascular response to stress:

Maximum heart rate after exercise. The maximum heart rate decreased significantly with age (Tab. 7). It is important to note that the maximum heart rates achieved in this exercise protocol were comparable to those in other studies, including those in which the maximum work load was significantly reduced with advanced age. That this result

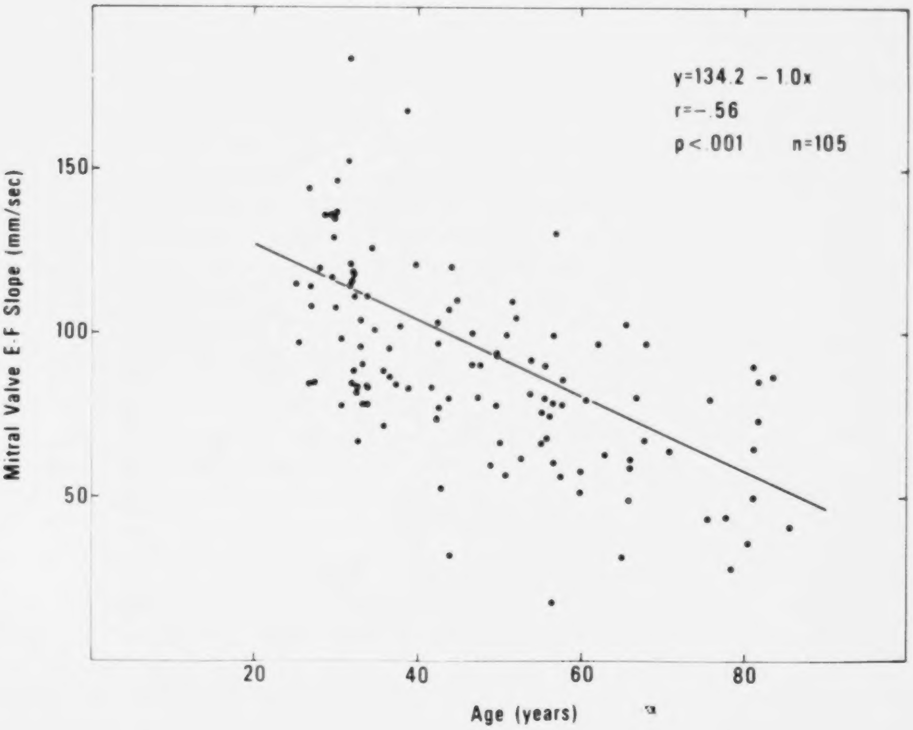


Figure V.18 Linear regression plot depicting the relationship between age and the rate of left-ventricular early diastolic filling as manifest in the closure rate of the anterior leaflet of the mitral valve (E–F) slope.
From Gerstenblith et al. (1977).

Table V.7. The Effect of Age on Hemodynamic Performance at Maximum Exercise in BLSA Subjects (N = 37)

Parameter	Linear Regression	r	p
Heart rate (BPM)	200.7 − .82 (age)	−.57	< .001
End-diastolic volume (cc)	126.2 + 1.28 (age)	.39	< .02
Stroke volume (cc)	107.6 + .69 (age)	.29	< .08
Cardiac output (4 min)	22.5 − .07 (age)	−.19	< .24
End-systolic volume (cc)	−.45 + .60 (age)	.50	< .002
Ejection fraction	92.9 − .20 (age)	−.35	< .03
Total peripheral vascular resistance	20.32 − .06 (age)	−.18	< .28
Systolic blood pressure (mm Hg)	177.9 + .34 (age)	.176	< .3

From Rodeheffer et al. (1981)

has been observed in virtually every population studied suggests that a diminished maximum heart rate is a true age-related phenomenon.

End-diastolic filling volume (pre-load). Blood filling the heart during diastole serves to stretch the fibers, in the process not only altering their geometry but also enhancing activation of the myofilaments. Through these mechanisms the heart can alter its stroke volume with changes in filling volume, an effect sometimes referred to as the Frank-Starling mechanism. It has previously been hypothesized (Gerstenblith et al., 1977) that because of changes in myocardial compliance—that is, the heart becomes stiffer with age—and because early diastolic filling rate is compromised with advancing age, diastolic filling volume may be lower in elderly than in younger adult subjects during exercise, when the filling time is reduced.

Not only is this not the case, but quite the opposite occurs: End-diastolic filling volume is increased rather than decreased as a function of age during the maximum exercise response (Tab. 7). In a previous study of another subset of the BLSA population at submaximal work loads—semi-supine exercise at a common heart rate of 120 bpm—the measured end-diastolic area determined by two-dimensional echocardiography was also greater in subjects 65+ years old than in those whose age averaged 30 years (VanTosh et al., 1980). The observed increase in filling volume was a major factor through which stroke volume not only was maintained, but actually tended to become larger, in the older group during exercise (Tab. 7). This in part balanced the effect of a decrease in maximal heart rate on maximal cardiac output, which in this population demonstrated only a slightly negative trend with age (Tab. 7). Enhancement of filling volume by the Frank-Starling mechanism may thus be construed as an age-related adaptive mechanism through which cardiac output is maintained during stress. A price is paid for this adaptation since, in accordance with LaPlace's law, generation of a given ventricular pressure requires greater wall stress (force per unit area) if the ventricular radius is increased; this in turn requires a greater level of cardiac work and energy production. Futhermore, enhanced filling volume, even in the absence of changes in compliance, results in an enhanced filling pressure, and may at least in part explain the age-related increase in filling pressure observed during exercise (Granath et al., 1964). An increase in left-heart diastolic filling pressure also produces an increase in pulmonary venous pressure, which increases the likelihood of pulmonary congestion.

Ejection fraction. An increase in diastolic filling volume should not only enhance stroke volume but also result in the ejection with each beat of a greater fraction of blood in the elderly than in young adults. The expected result, however, does not occur in the aged heart; end-systolic volume is not reduced during exercise in the elderly to the same extent as in younger adults; in fact, it increases with age (Tab. 7). Thus, although filling volume increases, the fraction of blood ejected is not increased but actually decreases with advancing age (Tab. 7). This indicates that the ejection of blood by the heart is compromised in the elderly subjects.

Autonomic modulation. During maximum exercise, when the sympathetic component is the exclusive autonomic modulator, a marked increase in catecholamine secretion occurs (Tzankoff et al., 1980). An age-related alteration in sympathetic modulation of the cardiovascular response to exercise could account for: a) the decline in maximum heart rate; and b) the apparent decline in maximum contractility, i.e., an increased end-systolic volume, and a decreased ejection fraction from a greater filling volume in the absence of an increase in systolic blood pressure in the elderly during exercise.

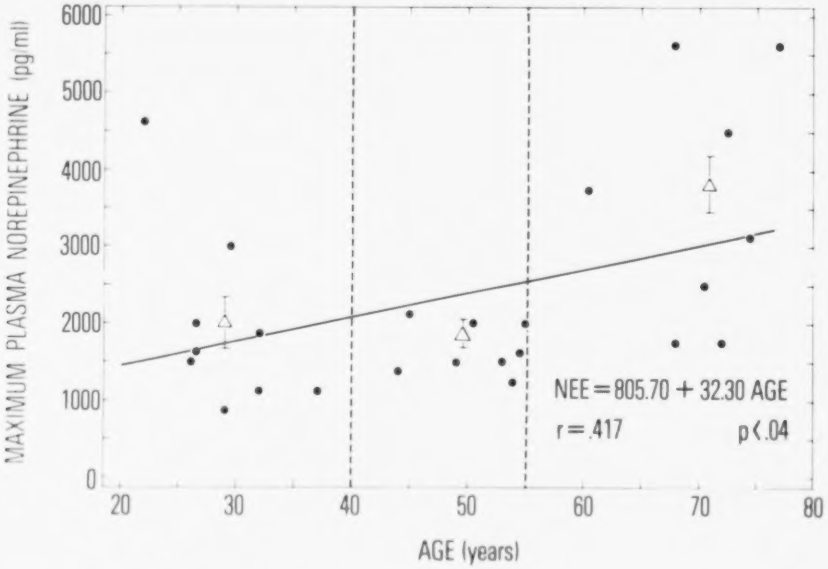


Figure V.19. The effect of age on plasma-norepinephrine concentration during maximum treadmill exercise in BLSA participants. Plasma norepinephrine concentration at rest did not differ in these subjects. In addition to the regression analysis across age, the subjects were divided into 3 age groups. The dotted vertical lines separate the 3 age groups in which the mean and SEM are denoted by △.
From Tzankoff et al. (1980).

A diminished sympathetic response during exercise could result from a deficiency in the elaboration of catecholamines in aged subjects. In a subset of the BLSA population (n = 27) both plasma epinephrine and norepinephrine, although not age-related at rest, demonstrated a substantial age-dependent increase during treadmill exercise (Fig. 19) (Tzankoff et al., 1980). A diminished maximum heart rate in these healthy subjects was thus accompanied by an increase rather than a decrease in plasma catecholamines. This result is consistent with the hypothesis that target-organ responsiveness to catecholamines declines with age. This was tested on a subset of the BLSA population by the infusion of incremental concentrations of isoproterenol in bolus form and monitoring of the resultant increase in heart rate (Lakatta, 1979). The same concentration of isoproterenol resulted in a greater increase in heart rate in younger than in older participants (Fig. 20). The result supports the notion that adrenergic responsiveness diminishes with age.

Summary: Hemodynamic Function at Rest and in Response to Stress

Age-related differences in the cardiovascular response to stress have often been observed in man. Their nature and magnitude have varied with the population studied. In apparently healthy populations that have not been screened for occult coronary disease, substantial decrements in maximum cardiopulmonary function and work capacity have been observed. In more selected populations, although maximum cardiovascular function does not markedly deteriorate with advanced age, definite age-related adaptations in hemodynamics serve to prevent substantial declines in cardiac

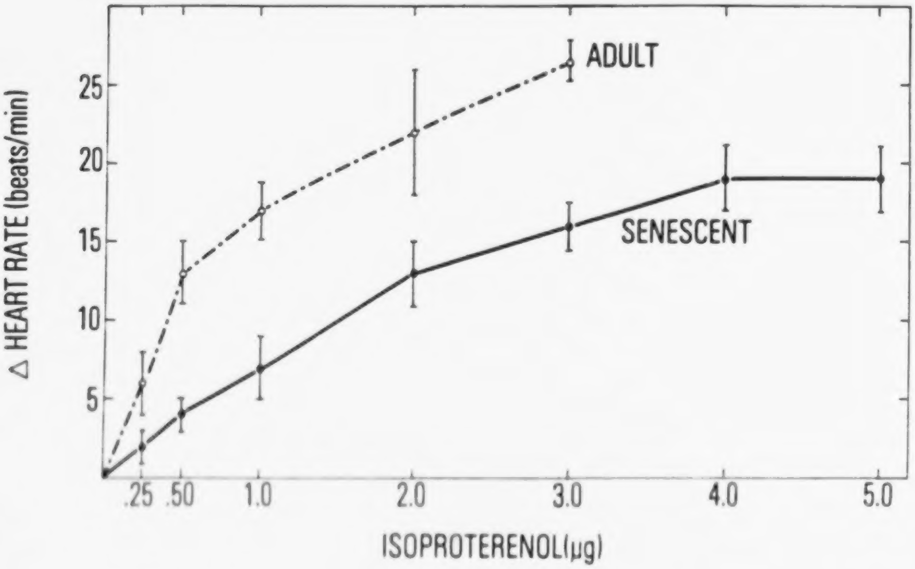


Figure V.20 The increase in heart rate induced by incremental intravenous boluses of isoproterenol in healthy adult (aged 18–34 yr) and 20 senescent (aged 62–80 yr) men. A significantly greater increment in heart rate is observed in the younger men at all isoproterenol dosages greater than 0.25 μg.
From Lakatta (1979).

output. A feature common to studies in a wide variety of populations is that the age-related differences observed may be attributed, at least in part, to altered responsiveness to adrenergic stimulation.

Electrophysiology

In addition to hemodynamic function, recent studies of BLSA participants have focused on the electrophysiological properties of the heart, in particular on the conduction system:

Non-invasive assessment of His-bundle ECG. It is not known whether the age-related increase in PR interval is due to slowed conduction proximal (PH) or distal (HV) to the His bundle, or to both. A microprocessor-assisted high-resolution ECG, which signal-averaged 512 cardiac cycles, recorded high-frequency, low-amplitude ECG signals from the body surface of 63 BLSA participants aged 21 to 70 ($\bar{x}$ age = 46.1 ± 17.0 yr) (Das et al., 1982). All subjects were clinically free of heart disease and displayed both normal resting ECG and normal maximal treadmill exercise tests. His-bundle potentials were identified in 53 individuals; the 10 persons in whom a His spike could not be identified did not differ in age or resting heart rate from those with His potentials, but had shorter PR intervals (131.5 ± 15.3 vs. 155.8 ± 18.5 msec, $p < .001$). In those subjects with demonstrable His activity, the following regressions were obtained:

$$PR = 139.6 \text{ msec} + 0.35 \text{ age}; \ (r = .32, \ p < .02)$$
$$PH = 101.1 \text{ msec} + 0.31 \text{ age}; \ (r = .33, \ p < .02)$$
$$HV = 38.3 \text{ msec} + 0.04 \text{ age}; \ (r = .09, \ p = NS)$$
$$HR = 72.6 \text{ beats/min} - 0.11 \text{ age}; \ (r = -.19, \ p = NS).$$

These observations led to the conclusions that: The success rate of recording surface His potentials varies directly with the PR interval; HV interval does not change with age in adults; and the prolongation of PR interval seen with advancing age is due to conduction delay proximal to the His bundle.

Ambulatory ECG in a healthy elderly population. Although recording of ambulatory 24-hour ECG has become a common diagnostic tool on which anti-arrhythmic and pacemaker therapy is predicated in the elderly, no suitable standards of normality exist for this age group. Ambulatory ECGs were analyzed in 98 BLSA subjects who had demonstrated a normal response to maximum treadmill ECG. Of this carefully screened population, 88% demonstrated supraventricular ectopic beats and 80% showed ventricular ectopic beats. A further breakdown of arrhythmias (Fleg and Kennedy, 1982) is shown below:

Supraventricular	*Percentage of Subjects*
≥ 100 beats in 24 hours	26
Benign slow atrial tachycardia	28
Paroxysmal atrial tachycardia	13
Ventricular	
≥ 100 beats in 24 hours	17
Multiform	35
Couples	11
Ventricular tachycardia	4

Sinus bradycardia less than 40/min and sinus pauses exceeding 1.5 seconds were rare. No instance of high-degree AV block or sinus arrest was observed. These results provide a framework for the interpretation of ambulatory ECG results in symptomatic elderly subjects.

7. Renal System

Creatinine clearance. Creatinine clearance is frequently used as a clinical estimate of kidney function (glomerular filtration rate, GFR). Although the published criteria for severity of renal disease included classification of patients according to their creatinine

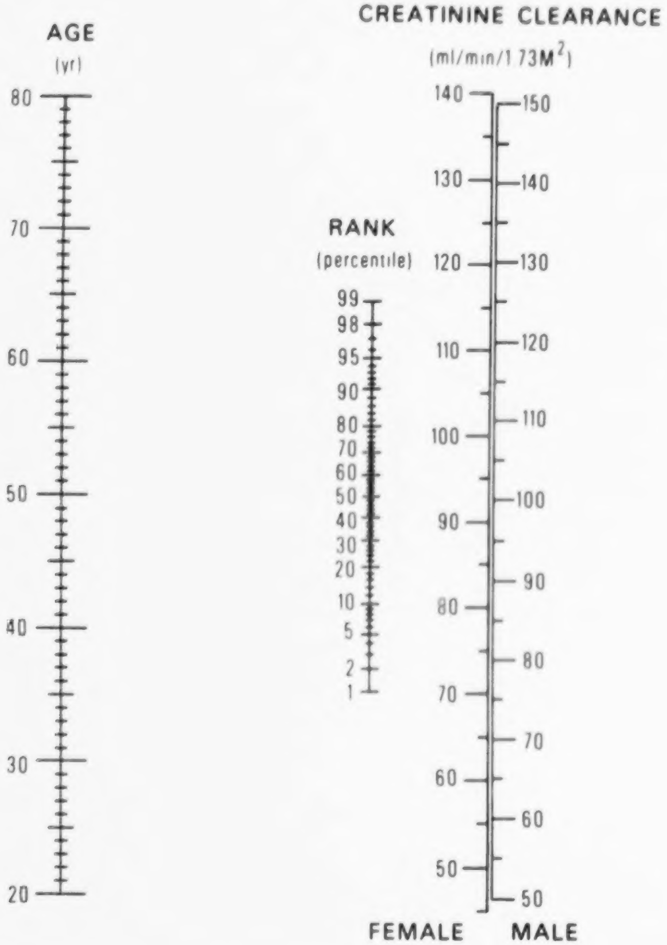


Figure V.21. Nomogram for ascertaining age-adjusted percentile rank in creatinine clearance. For use with creatinine determinations by automated "total chromogen" method. See text for derivation of data for females. A line through the subject's age and creatinine clearance intersects the percentile rank line at a point indicating the subject's age-adjusted percentile rank. From Rowe et al. (1976a).

clearance in comparison with the "predicted normal," no age-adjusted normal standards for creatinine clearance were available until 1976, when the GRC provided them (Rowe et al., 1976a). Standard 24-hour creatinine clearance measurements were made on 884 adult male participants (age range = 17–96 yr) in the BLSA. True creatinine was measured in both serum and urine. There was a highly significant reduction in creatinine clearance with age in 548 carefully screened subjects, and the relation between creatinine clearance and age was found to be best described by the linear equation: Creatinine clearance ($\text{ml/min} \cdot 1.73\text{m}^2$) = $133 - 0.64 \cdot \text{age (yr)}$. Nomograms for ascertaining age-adjusted percentile rank in creatinine clearance were constructed for men and, by derivation, for women (Fig. 21). The conversion factor of 0.93 to derive values for females was taken from the literature review by Wesson (1969). The longitudinal analysis of renal function is summarized in Chapter VI, and the study *(Rowe et al., 1976b)* is reprinted in the Appendix.

Concentrating ability. Several cross-sectional studies have shown that the ability to concentrate urine declines with age after maturity. A BLSA study was performed to determine cross-sectional age differences in solute and water conservation, independent of the effects of disease or medication (Rowe et al., 1976c). The 97 participants included in the analysis ranged in age from 20 to 79 years. At 6:00 P.M. each subject was asked to void and the urine was discarded. Subjects were then instructed not to eat or drink until 6:00 A.M. Urine specimens representing three time periods were then collected between 9:00 P.M. and 6:00 A.M. At 6:00 A.M. a blood sample was collected for determination of osmolality and creatinine concentration. At 10:00 A.M., a final urine sample was collected, and the total 16-hour pool was used for determination of creatinine clearance. During Period 1 (6:00-9:00 P.M.), no significant age-related differences were found in urine osmolality, urine flow, solute secretion, or osmolar clearance. By Period 3 (midnight to 6:00 A.M.), however, young subjects had responded to water deprivation with a marked decrease in urine flow, a moderate increase in urine osmolality, and a resultant significant decrease in osmolar clearance. Middle-aged subjects also decreased urine flow and increased urine osmolality, but to a lesser extent than young subjects; the result was a small net decline in osmolar clearance. In the elderly subjects urine-flow rate, urine osmolality, and osmolar clearance were not significantly altered during the period of dehydration. These effects indicate that in addition to the reduction in glomerular filtration rate there is an impaired response of renal tubules to changes in plasma osmolality in the old subjects.

8. Endocrine System

Age and diabetes. The cortisone glucose tolerance test, introduced in the mid-1950s, was proposed for the detection of the prediabetic state, that is, to provide a more sensitive index of diabetes than the oral glucose tolerance test. Several studies had shown that poor performance on the cortisone glucose tolerance test was predictive of the future development of clinical diabetes.

Although it had been shown by 1965 that performance on both the standard glucose tolerance test and the cortisone glucose tolerance test declines with age, the studies failed to quantify the age effect or define the age at which it begins (Andres, 1971). To test the assumption that the age-related loss of glucose tolerance might be clinically significant, a BLSA study was undertaken to quantify the effect of aging on cortisone glucose tolerance test performance over the entire adult age span (Pozefsky et al., 1965).

Of the 89 male subjects, aged 21 to 95 years, initially selected for study, data from 15 were eliminated because they had diabetes, reported a positive family history of diabetes that was less than first-degree, or were taking drugs known to alter carbohydrate metabolism. An additional 13 subjects with a primary family history of diabetes were analyzed separately. The final group consisted of 61 individuals.

Cortisone acetate based on body weight was given orally in two equal doses 8.5 and 2 hours before glucose administration. After withdrawal of a fasting blood sample, 1.75 g of glucose/kg of body weight was administered orally as a 30% solution. Blood samples were then obtained at 20-minute intervals for two hours. Blood glucose was measured by ferricyanide reduction. An obesity index was computed for each subject from the ratio of actual weight to the mid-weight for the medium frame in the 1959 Metropolitan Desirable Weight Table (Metropolitan Life Insurance Co., 1959).

Fasting blood-glucose levels were elevated by cortisone in subjects of all ages, but sensitivity to this effect increased significantly with advancing age: The average increment was 2.1 mg/dl for each age decade. Blood-glucose concentration also increased with age after glucose administration. By 120 minutes the average increase in glucose level was 17.6 mg/dl per decade of life. The subjects were divided into three age groups (20–44, 45–64, and 65–95 yr) for comparison of mean cortisone glucose

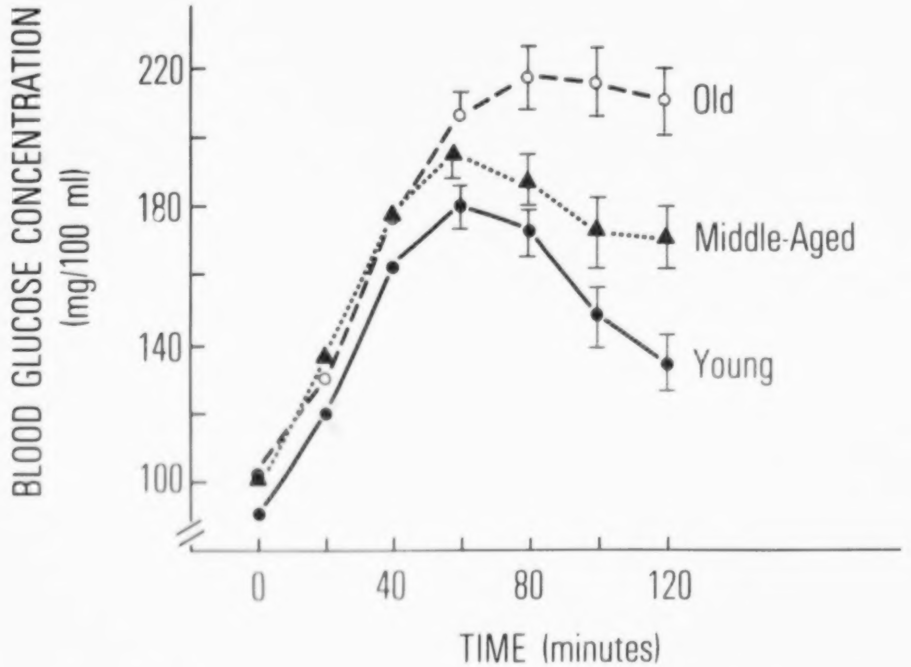


Figure V.22. Mean cortisone-glucose tolerance test curves for negative-family-history subjects in 3 age groups. The SEM for each time period after 40 min is also shown. Differences between these groups were not significant before 60 min. The 15 young subjects were aged 21-44, the 26 middle-aged 45-64, and the 20 old 65-95 yr.
From Pozefsky et al. (1965).

tolerance test curves. The oldest group reached peak blood glucose levels at 80 minutes and showed only a slight decline thereafter. Young subjects peaked at 60 minutes, and fell more rapidly in the second hour of the test. The response of the middle-aged group was intermediate (Fig. 22). The effect of age on glucose levels was thus not simply a difference between young and old but manifest throughout the adult age range.

A comparison of tolerance curves revealed higher glucose intolerance in subjects whose families had histories of diabetes. The differences between this group and the control group, which although small were statistically significant, confirmed the observation that performance on this test is poorer in subjects at higher risk of developing diabetes.

The profound age influence on performance made it clear that reliance on a single standard for the upper limit of normality was unrealistic. A nomogram was therefore proposed to permit percentile ranking of an individual within his age group, although prospective studies are required to determine the prognostic implications of specific percentile ranks (Fig. 23).

The application of separate standards to different age groups assumes that deterioration in glucose tolerance with increasing age does not necessarily indicate diabetes mellitus. The question whether the decrement is the result of pathology or a

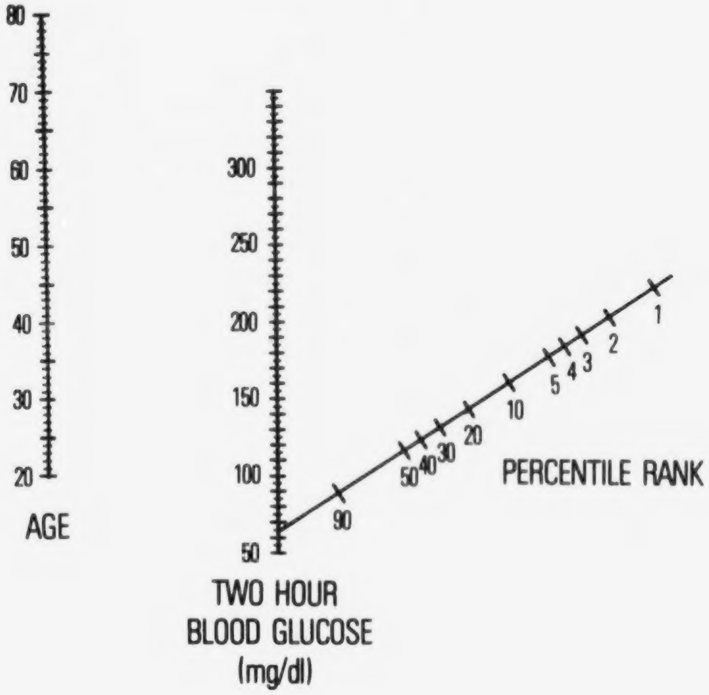


Figure V.23. Cortisone-glucose tolerance test. Nomogram for establishing an individual's percentile rank in his birth cohort, based on 2-hr blood-glucose concentration.
From Pozefsky et al. (1965, adapted).

part of normal aging might be solved by a test for diabetes that does not involve measuring glucose tolerance. A BLSA study (Swerdloff et al., 1967) was undertaken to test the possibility that the intravenous tolbutamide test might be an age-independent method of detecting diabetes without reference to glucose tolerance. The decline of blood-glucose concentration that occurs after tolbutamide administration is markedly influenced by diabetes, occurring more slowly and to a lesser degree than in normal individuals.

Tolbutamide tests were performed in 141 male subjects, 117 of whom (age range = 25–81 yr) were randomly selected from participants in the BLSA. The other 24 were college students and hospital employees 21 to 36 years of age. Forty-one subjects were

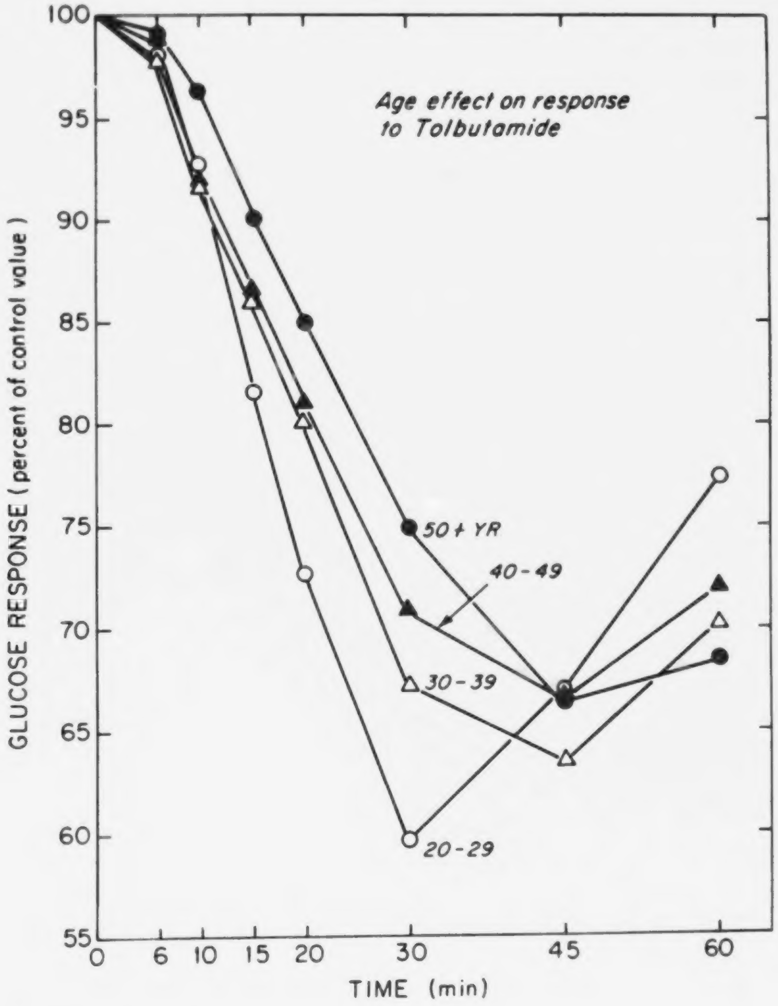


Figure V.24. Age effect on response to tolbutamide. From Swerdloff et al. (1967).

subsequently excluded for characteristics that were likely to confound the results, such as those noted above for the cortisone-glucose tolerance test. After overnight fasting, a baseline blood sample was taken. Then sodium tolbutamide in distilled water at a dose of 1 g/70 kg body weight was injected, and blood samples for glucose analysis were obtained at 6, 10, 15, 20, 30, 45, and 60 minutes.

A significant increase in fasting blood-glucose concentration with advancing age was found in the group, in contrast with other groups in which there was no difference with age, but the increment was only 0.9 mg/dl per decade of life. The fall in glucose concentration after tolbutamide administration was markedly affected by the age of the subject. Blood-glucose levels fell rapidly in the youngest subjects, usually reached a minimum in 30 minutes, and generally rebounded by 60 minutes. In the older subjects, glucose fell more slowly, reaching a minimum at an average of 45 minutes, and the decrease was smaller. The greatest age effects were seen at 20 and 30 minutes after tolbutamide administration (Fig. 24). The decline in response to tolbutamide was progressive in the four decades from age 20 to age 59. There was, however, no further decline in response in subjects over age 59.

Since age effects were pronounced in the tolbutamide test, it was concluded that this method, like the glucose tolerance method, requires age adjustment for proper interpretation. Accordingly, a nomogram was provided to permit the rapid computation of age-adjusted percentile ranking of an individual's response to intravenous tolbutamide (Fig. 25).

The limitations of various methods of diagnosing diabetes in the elderly were discussed by Andres (1971) in a symposium on diabetes mellitus. He concluded that statistical manipulations of survey data cannot resolve the question whether standards of normality should be adjusted for age, and recommended that long-term longitudinal follow-up studies be made in populations that have participated in glucose tolerance tests, to determine what levels of performance in different age groups indicate increased risk of diabetes and such diabetes-related events as atherosclerosis, diabetic neuropathy, nephropathy, retinopathy, or, indeed, death itself. Once a high-risk group is identified, he added, the efficacy of therapeutic regimes can be assessed.

Another BLSA study (McGuire et al., 1979) provided some insight into the changes in insulin kinetics *in vivo* that are characteristic of aging, of moderate obesity, and of maturity-onset diabetes insufficiently severe to require insulin therapy. The kinetics of unlabeled porcine insulin were studied in 69 nondiabetic males aged 18 to 83 years with obesity indexes of 0.93–1.51, and in 12 maturity-onset diabetics aged 46 to 78 years with obesity indexes of 0.95–1.56, by use of the insulin-infusion glucose-clamp technique (DeFronzo et al., 1979) to maintain constant blood-glucose levels. The individuals were grouped to allow comparison of the results on the basis of age, obesity index, or diabetes. Analysis of the kinetic data with a mathematical model allowed steady-state distribution masses and degradation-rate constants to be determined for each subject.

The responses over a period of 120 minutes to an intravenous infusion and washout of insulin showed both transient and steady-state differences with age, obesity, and diabetes. Analysis of the data led to the conclusion that in the steady state the ratio of insulin in extravascular spaces to insulin in plasma was decreased by 26% in the moderately obese group and by 17% in the diabetic group, but was increased by 13% in the older group, when each was compared with the appropriate control. It was concluded that these changes probably reflect changes in the binding of insulin to receptors, although the magnitude of the changes would be somewhat modified by alterations in the size of the interstitial space in relation to plasma volume.

In addition, the rate of entry of new insulin into plasma was increased by 45% in the diabetics and by 27% in the moderately obese group, but was decreased by 11% in the older group.

These findings led to the following general conclusions. The pattern of changes seen with obesity is similar to that in maturity-onset diabetes. In obesity and maturity-onset diabetes, the decrease in the ratio of insulin in extravascular spaces to insulin in plasma cannot be accounted for solely by changes in fasting plasma-insulin levels. The fact that the pattern of changes seen in the older subjects is the opposite of that in maturity-onset diabetics suggests that diabetes is distinct from the changes in normal aging. Finally, since the changes in the metabolism of insulin are small, it is unlikely that they are the sole cause of the major alterations in glucose tolerance that occur with age, obesity, or diabetes.

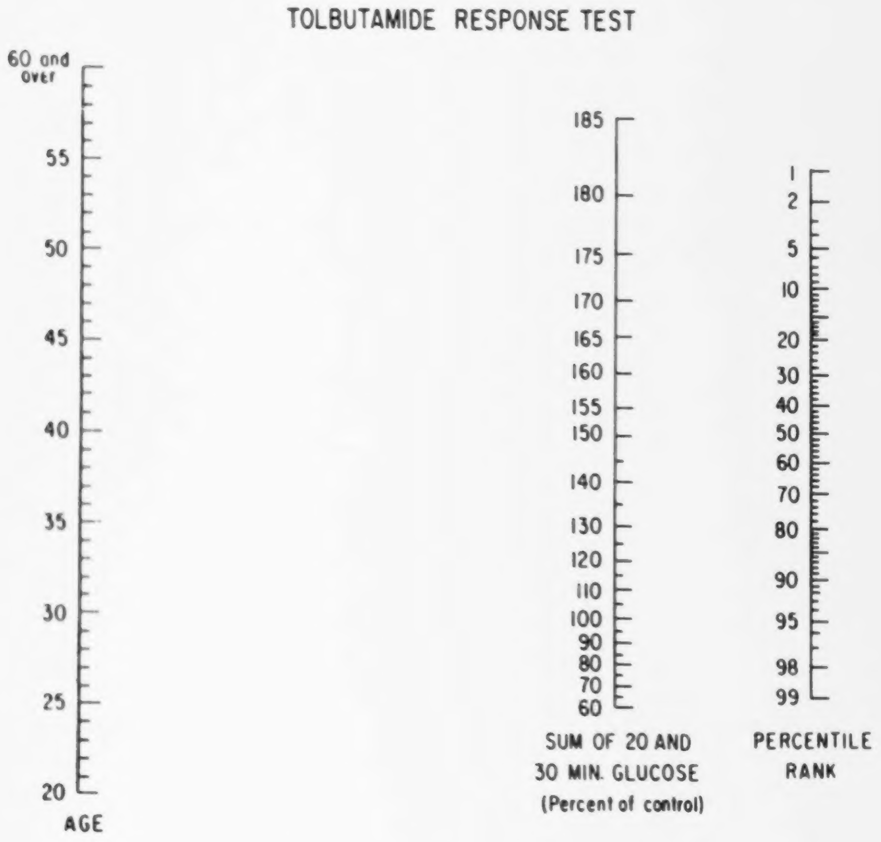


Figure V.25. Nomogram for judging percentile rank on tolbutamide response test using sum of 20- and 30-min values.
From Swerdloff et al. (1967, adapted).

Hypothalamic-pituitary-testicular axis. A number of investigators looking at the effects of age on sex hormones in men have come to the conclusion that the major male hormone, testosterone, decreases with age (Stearns et al., 1974; Vermeulen et al., 1972) and that, because of an increase in the specific sex hormone binding globulin (SHBG), the unbound (bioavailable) hormone decreases even more prominently than the total. Along with these changes, increases in plasma estrogens have also been reported (Rubens et al., 1974; Pirke and Doerr, 1975). Most investigators have found an increase with age in the plasma levels of the pituitary gonadotropins, luteinizing hormone (LH), and follicle-stimulating hormone (FSH); the increase presumably reflects the decline in testicular function with subsequent reduction in inhibition of gonadotropin secretion by negative feedback (Stearns et al., 1974; Vermeulen et al., 1972; Baker et al., 1976).

In a study of 76 BLSA men (Harman and Tsitouras, 1980), we measured basal circulating levels of gonadotropins, four sex steroids (testosterone [T], dihydrotesterone [DHT], estradiol [E_2], and estrone [E_1]), an index of sex steroid binding to plasma globulin, testicular volumes, and, when possible, sperm production. An index of sexual activity was obtained by Dr. Clyde Martin of the Human Performance Section. In addition, provocative studies of pituitary and gonadal secretory function used luteinizing hormone-releasing hormone (LRH) (Harman et al., 1982) and human chorionic gonadotropin (hCG) to test secretory reserve capacity. Results of these studies can be summarized as follows:

Unlike the subjects in previous investigations, who showed decreased T and free T index and increased E_1 and E_2, our healthy volunteers revealed no significant changes in prevailing blood levels of sex steroids over the age range from 25 to 90 years.

Leydig-cell reserve capacity was somewhat diminished after hCG stimulation, although the differential narrowed with time after hCG stimulation.

Basal levels of both LH and FSH were increased with age despite the absence of a change in sex steroids.

Pituitary response to LRH:

- LH response, after correction for effect of basal LH level, was moderately decreased in the oldest men (age range = 70–89 yrs), but significantly so only in comparison with the middle-aged men (age range = 50–69 yr). LH peak response time was significantly delayed with age.
- FSH response showed a significant decrease with age, but no delay in peak time.
- Glycoprotein hormone sub-unit secretion. Measurements of α and LHβ and FSHβ sub-units in the sera obtained in this study showed increases in the basal levels of each sub-unit with age, but these increases were low in proportion to the increase in intact gonadotropins. Thus, the ratios of α/(LH + FSH) decrease as gonadotropins rise. This phenomenon presumably results from more efficient synthesis and combination of sub-units into the heterodimer form. Responses of α, LHβ, and FSHβ to LRH injection were reduced in aged men (after correction for basal level), but to a lesser extent than the LH and FSH responses. The secretory ratios of sub-unit to intact hormone after LRH stimulation were thus relatively higher in aged men than in their younger counterparts.

In unpublished data sexual-activity levels decreased with age, despite the maintenance of normal sex-steroid levels. Nonetheless, high free T index was weakly but significantly correlated with higher sexual-activity level, and low free T index with reduced activity in the oldest (and only in the oldest) men. No significant differences were found in testicular volumes with age, nor in mean sperm counts per ml or per ejaculate.

From the above data, we conclude that elderly but healthy middle-class men, living at home, may not manifest decreased androgens or increased estrogens. These findings contrast with those in previous study populations, which often included older subjects from clinics, nursing homes, and similar institutions and may thus have reflected the effects of variables other than aging *per se*. The reduced hCG response and elevated gonadotropin levels suggest that although some age-related defect in testicular function may occur in these men it is compensated by increased pituitary activity. The pituitary-response data suggest reduced sensitivity of pituitary gonadotrophs to LRH with age. The decreased sexual activity in men with normal androgens must be produced by something other than hypogonadism, perhaps by alterations in central or peripheral nervous-system function. The better-maintained sexual function in old men with higher free T indices suggests that high levels of androgen activity may partially overcome the age-related deterioration of sexual function.

In order further to investigate the relationships of aging, sexual activity, and sex hormone levels, we used lyophilized sera and other data obtained from 180 BLSA men aged 60 to 80 years (Tsitouras et al., 1982) in a retrospective survey of the relation of sexual activity to serum T, body composition, alcohol and tobacco habits, and cardiac status. The study confirmed the correlation of serum T and sexual-activity level in older men and suggested that the relation is not mediated by obesity, muscle mass, alcohol consumption, smoking, or coronary artery disease. A tendency toward decreased sexual activity, but no decrease in T, was found in men consuming the largest amounts of alcohol, and a tendency toward decreased T, but no decrease in sexual activity, in men having the greatest percentages of body fat.

Arginine vasopressin. Secretion of arginine vasopressin (AVP), a natural peptide antidiuretic hormone, by the hypothalamic-neurohypophyseal-renal axis is inhibited by ethanol and stimulated by hypertonic saline. The impact of age on these responses was determined by direct assay of AVP (Helderman et al., 1978). Subjects in two age groups (9 men 21-49 and 13 men 54-92 yr) were administered ethanol intravenously at the rate of 375 mg/m^2 surface area per minute for one hour. Sixteen other subjects in two age groups (age ranges = 22-48 and 52-66 yr) were given an intravenous infusion of 3% NaCl for two hours at the rate of 0.1 ml/kg body weight per minute. Young and old control subjects received normal saline intravenously. Blood and urine samples were collected from all subjects and analyzed. AVP levels fell progressively during ethanol infusion in the younger group, but in the older group they fell for only 30 minutes, then paradoxically rose to nearly basal levels as ethanol levels continued to increase; after infusion ceased, the AVP values continued to rise to nearly twice basal values. In the hypertonic saline group, serum AVP rose to 2.5 times the basal level in young men and 4.5 times the basal level in old men despite identical free-water clearances. The paradoxical AVP response of the older subjects to greater osmolality induced by ethanol administration was partly due to their significantly greater osmoreceptor sensitivity (Fig. 26), which may compensate for the decreased ability of their aging kidneys to conserve salt and water.

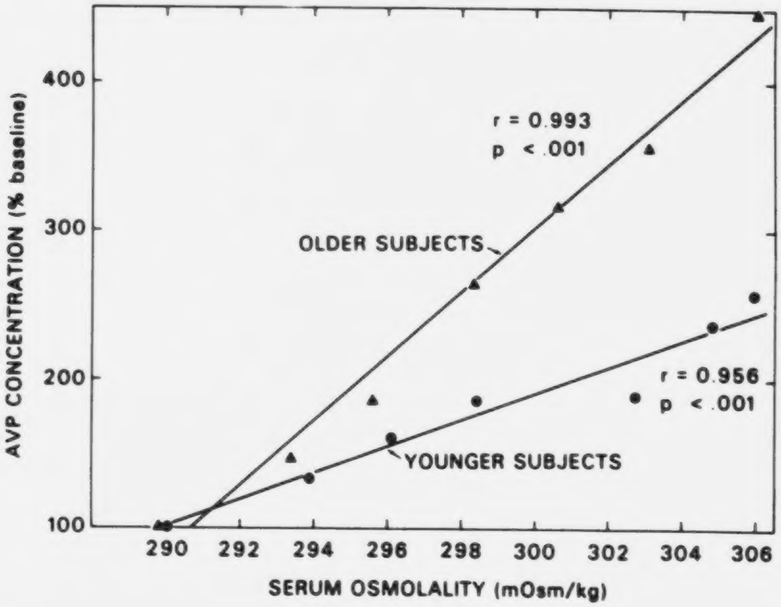


Figure V.26. Correlation between serum osmolality and AVP concentration in young and old subjects during 2-hr intravenous infusion of 3% NaCl at the rate of 0.1 ml/kg per min. The points represent mean values of osmolality and AVP at successive 20-min intervals in each group. From Helderman et al. (1978).

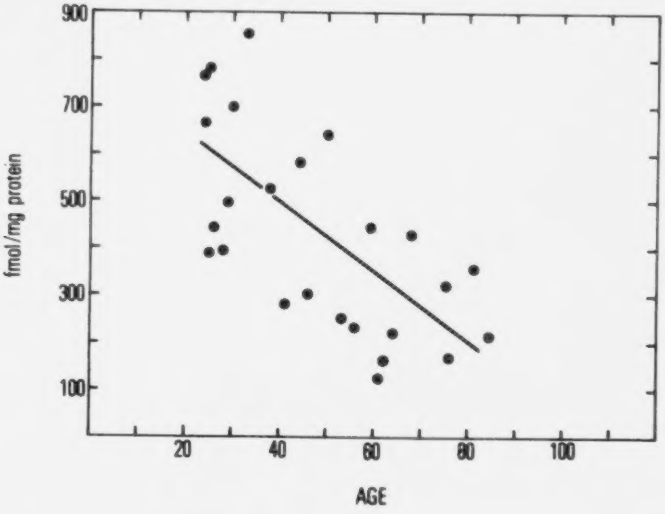


Figure V.27. Maximal specific binding of (——) ^{3}H-dihydroalprenolol to crude mononuclear cell membranes of subjects aged 24–81 yr.
From Schocken and Roth (1977).

Hormone receptors. Although previous studies have shown that the concentrations of hormone receptors are reduced during aging in several animal species, most of the work has been limited to intracellular receptors for steroid hormones. A BLSA study (Schocken and Roth, 1977) found that cell-surface binding sites for β-adrenergic hormones in humans also decrease in number with advancing age. Mononuclear cell fractions obtained from BLSA subjects aged 24 to 81 years were assayed for β-adrenergic receptors using ^{3}H-dihydroalprenolol. Although the binding sites retained a high affinity for dihydroalprenolol irrespective of the age of the donor, the concentration of receptors dropped significantly with age (Fig. 27). The mean saturation level was 572 ± 58 pmol/mg protein in subjects 24 to 41 years old and 332 ± 48 pmol/mg protein, or slightly more than half, in subjects over 46 years. Likewise, the number of sites per cell correlated inversely with age (14,000 in the young and 8000 in the old group). While previous hormone-receptor studies in man have dealt with either intracellular receptors or cells in tissue culture, this investigation was the first to demonstrate age-associated alterations of surface hormone binding sites in cells taken directly from humans. The results support the idea that loss of certain hormone receptors is a common manifestation of aging and also offer a possible explanation of age-associated loss of responsiveness to adrenergic hormones.

More recently, a β-adrenergic binding site with an affinity approximately ten times higher has been detected by two other laboratories (Abrass and Scarpace, 1981; Landmann et al., 1981). Its concentration is apparently not affected by aging, although β-adrenergic stimulation of adenylate cyclase is markedly reduced in lymphocytes of aged subjects. Thus β-adrenergic mechanisms in these cells may be altered at the level of adenylate cyclase itself, and the role of age changes in the lower-affinity binding site remains to be elucidated.

It has been suggested that the insulin binding system undergoes a genetically programmed aging. Previously reported differences in insulin binding in fibroblast samples are most striking when data from children are compared with data from adults. Since the differences may be interpreted as representing developmental changes rather than adult aging, a study (Hollenberg and Schneider, 1979) compared ligand binding and biological responsiveness to porcine insulin and murine epidermal growth factor-urogastrone (EGF-URO) in skin fibroblast cultures derived from men 22 to 31 years old with those of cultures from men 65 to 80 years old. The receptor characteristics of the cells from the two age groups did not differ.

9. Immune System

Immune function has so many facets that it requires a variety of assay procedures, including *in-vitro* cell culture and the determination of serum immunoprotein and antibody levels. Although serum samples can be obtained from all BLSA subjects during their visits to the GRC, the culture of the peripheral blood white cells and some of the cell-culture assays are so time-consuming that cells from only one individual can be examined during a day. As a result of the difference in time and effort required by various assays, the number of individuals who constitute the data base varies with the assay, and not every assay has been performed on every BLSA participant. The first set of assay results to be discussed is concerned with cellular function in the host immune response.

Number of white cells in peripheral blood. From longitudinal data collected during five visit cycles, the number of the various types of white cells present in the peripheral

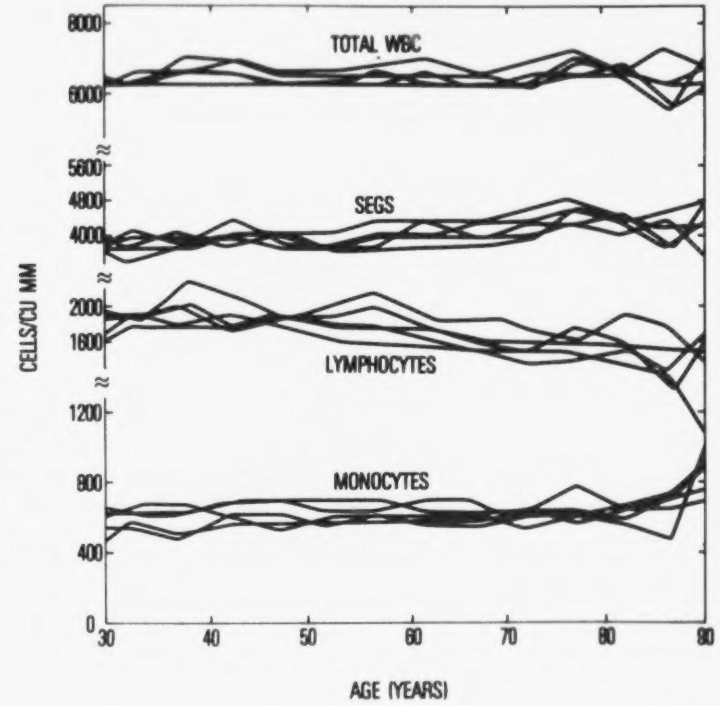


Figure V.28. Counts of total white blood cells (WBC), segmented neutrophiles (SEGS), lymphocytes, and monocytes were determined at each visit of the BLSA participant. Visits occurred at 1- or 2-yr intervals. Each line in each grouping represents one cycle of visits. Five cycles are presented for each white-cell group.
From Adler and Nagel (1981).

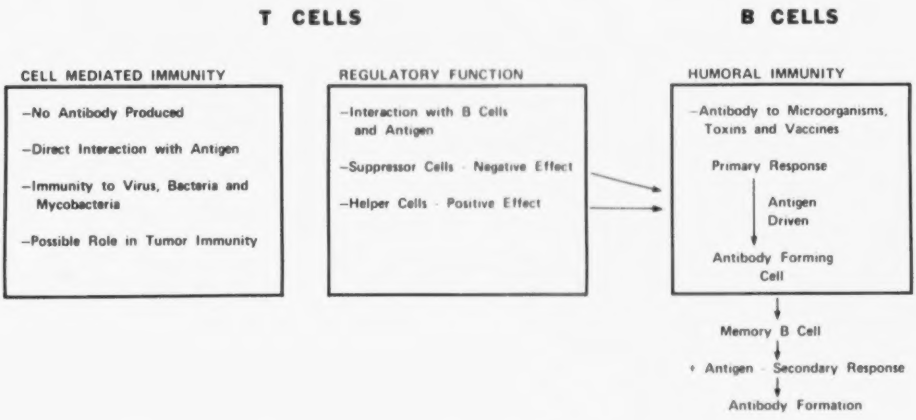

T CELLS
B CELLS
CELL MEDIATED IMMUNITY
–No Antibody Produced
–Direct Interaction with Antigen
–Immunity to Virus, Bacteria and Mycobacteria
–Possible Role in Tumor Immunity
REGULATORY FUNCTION
–Interaction with B Cells and Antigen
–Suppressor Cells - Negative Effect
–Helper Cells - Positive Effect
HUMORAL IMMUNITY
–Antibody to Microorganisms, Toxins and Vaccines
Primary Response
Antigen Driven
Antibody Forming Cell
Memory B Cell
+ Antigen - Secondary Response
Antibody Formation

Figure V.29. Functional characteristics of T lymphocytes and B lymphocytes.

blood was determined over the age range from 20 to 90 years (Fig. 28). There were no differences in the total white-cell count or in the number of the various white cells in the peripheral blood (Adler and Nagel, 1981; Nagel et al., 1982b). Granulocytes, lymphocytes, and monocytes maintain the same relative representation and absolute numbers throughout this age range. Use of antisera with specificity against membrane components makes it possible to examine subtypes of lymphocytes such as the B lymphocyte, which has antibody-synthesizing ability, and the T lymphocyte, which is thymic-derived and important in host defense in cell-mediated immune activity and serves in the regulation of B-cell activity (Fig. 29). Antisera with binding specificity against B cells, T cells, and the subsets of T cells are available to quantify these populations (Nagel et al., 1981b; Nagel et al., 1983). Figure 30 shows that the representation of B cells in the peripheral blood of individuals of different ages remains constant. The determination of T cells and T-cell subsets has been studied not across the age span but in young (age range = 20–40 yr) and old (60+ yr) subjects. It was found that the total number of T lymphocytes is lower in the older population because

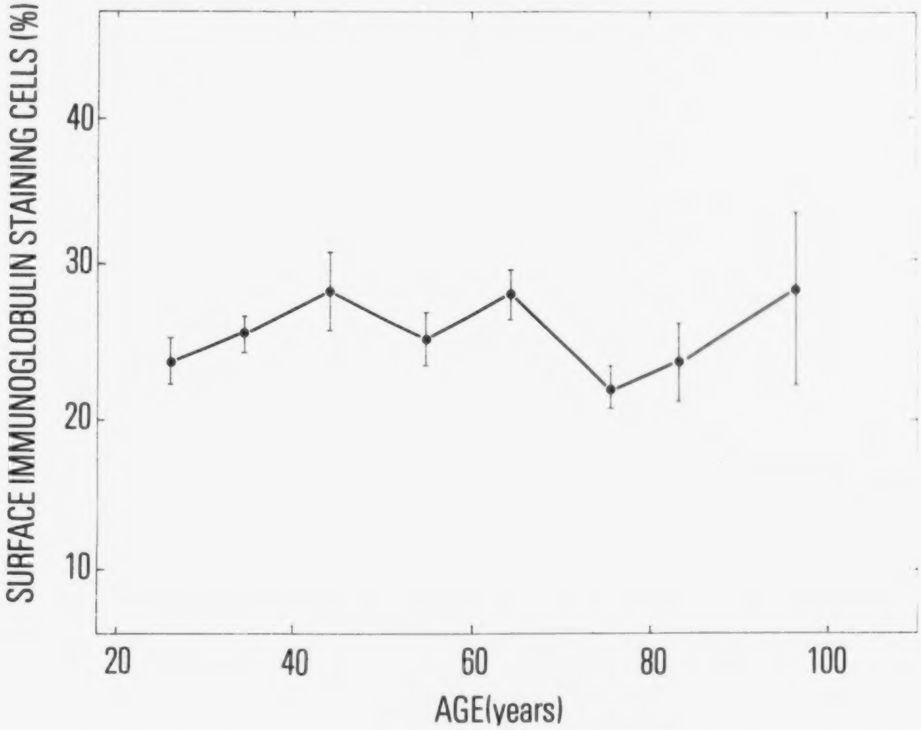


Figure V.30. The percentage of blood lymphocytes with membrane immunoglobulin was determined on 350 individuals of varying age. Each point represents the mean value for a 5-year interval and is plotted in the mean position for that age group. The bar at each point represents plus or minus one SEM.
From Adler and Nagel (1981).

160

of the decrease in the T-cell subset known as the "cytotoxic-suppressor" population (Fig. 31). These studies led to the *in-vitro* cell-culture experiments detailed below.

Granulocyte function tests. One arm of the host defense against infectious disease is the function of the polymorphoneutrophile (PMN). Since previous quantitative morphologic work had shown no change with age in PMN numbers in peripheral blood, it was important to determine PMN function. A variety of assays were used. Metabolic activity was measured by Nitro-Blue-Tetrazolium (NBT) dye reduction. The reduction of NBT to formazan requires the generation of H$^+$ through the metabolism of glucose. Another system measured the generation by the PMN, in response to a

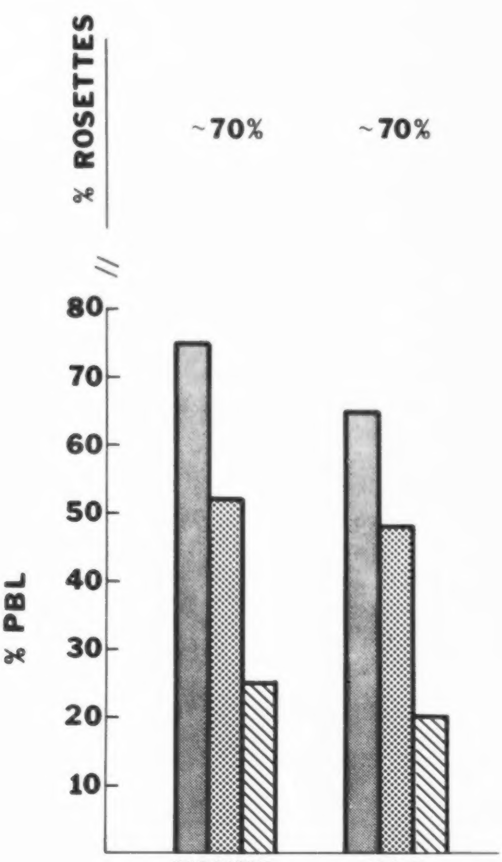


Figure V.31. The percentage of blood lymphocytes that react with the monoclonal antibodies OKT3 (an anti pan T cell antibody) ■ , OKT4 (an anti T helper cell) ▦ , or OKT8 (an anti T suppressor, cytotoxic cell) ▧ was determined on samples from young (<age 40) and old (>age 60) BLSA participants. Rosette determination for T-cell identification showed no age-related changes. The monoclonal antibodies showed a drop in the number of T cells in the old group that was due mostly to a decrease in the suppressor population.

stimulus, of lysosomal enzyme, which is necessary for the PMN to kill ingested bacteria. The last assay measured phagocytic and bacteria-killing ability of the PMN challenged *in vitro* by staphylococci. None of the assays demonstrated age-related change in granulocyte PMN function *in vitro* (Nagel et al., 1982b) (Figs. 32, 33). The study has not, however, addressed the question of PMN function and reserves in the whole person, especially during a period of physiologic stress.

Antibody-forming ability of peripheral blood lymphocytes (PBL). Since morphologic studies of the PBL subpopulation had shown relatively minor age-related changes, a study was undertaken to measure the functional ability of these cells *in vitro*. PBL from

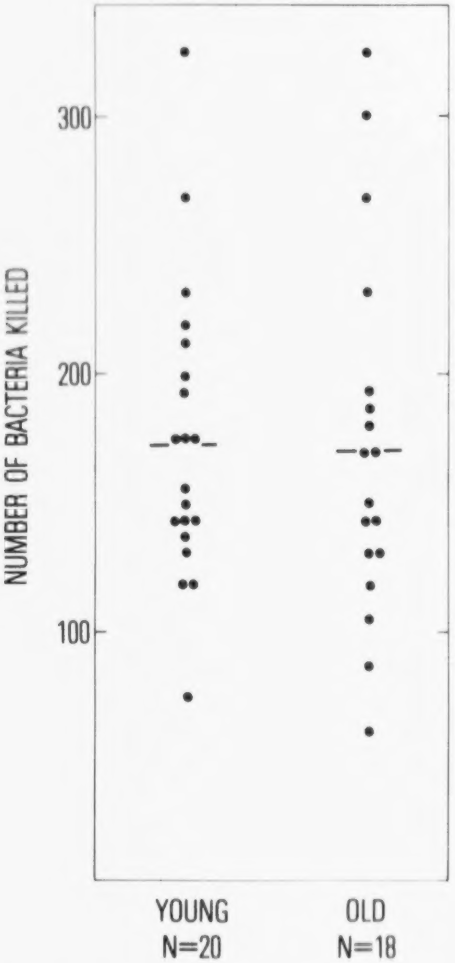


Figure V.32. The bactericidal capacity of granulocytes was determined *in vitro* using the *Staphylococcus aureus* organism. Young participants were less than than 40 years of age, old over 60. The mean number of bacteria killed by PMNs in 60 min was the same in both groups. From Nagel et al. (1982b).

individuals were placed in culture with a polyclonal B-cell activator, Pokeweed Mitogen (PWM). PWM will induce antibody synthesis in B cells if the appropriate functionally active helper T-cell population is present. In a large series of determinations of antibody synthesis *in vitro*, it was found that PBL in about 13% of the BLSA individuals, most of whom were over age 50 (Fig. 34), could not be induced to synthesize immunoprotein (Nagel et al., 1981a). More detailed analysis of cells from the nonresponder group demonstrated that B-cell function was normal and that the lack of response resulted from a decreased ability of the helper T-cell population to interact with the B cell (Chrest et al., 1983) (Fig. 35). Without a functional helper T cell, the B cell cannot be induced to synthesize antibody protein.

Natural killer (NK) cell activity. A population of lymphoid cells found in human peripheral blood lacks the membrane markers associated with T or B cells. These, designated "null" cells, have an interesting *in-vitro* function that is indicated by their

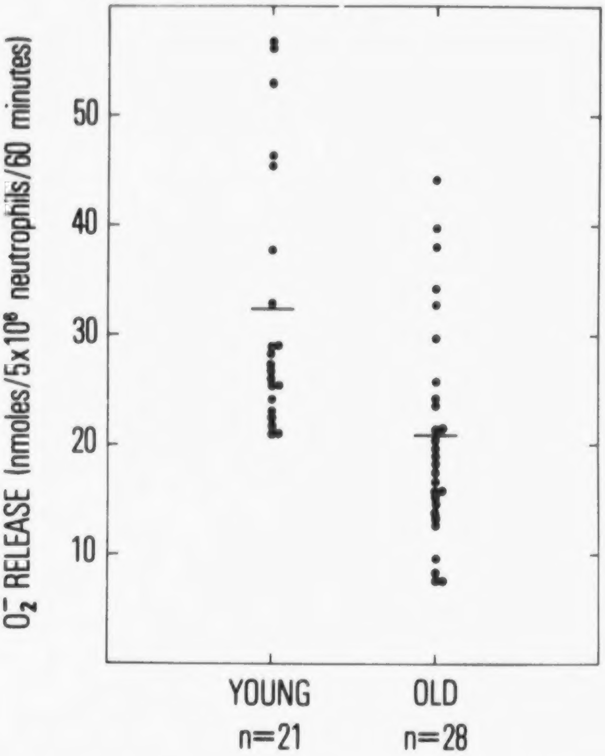


Figure V.33. Superoxide generation by granulocytes from young (<age 40) and old (>age 60) was determined in stimulated PMN cultures. Although the mean superoxide level generated by PMNs from elderly participants was lower, there was no significant difference between the groups.
From Nagel et al. (1982b).

further designation as "natural killer" cells: They can kill a variety of tumor cells *in vitro* without previous sensitization with the tumor-cell antigen either *in vivo* or *in vitro*. A survey of the BLSA volunteers for levels of NK activity in their PBL populations revealed no age-related change (Nagel et al., 1981c). Nor was there any difference between the male and female groups (Fig. 36). Another finding was that NK-activity levels of PBL from the same individual remained relatively the same over a three-year period. A dietary factor, the ingestion of ethanol (Saxena et al., 1980), was nevertheless found to elevate NK activity. The NK activity of PBL in individuals who regularly consumed ethanol was consistently higher than that in age-sex-matched controls. A dietary factor can thus change a result in an assay system. But since the significance of NK-cell activity *in vivo* is not now known, it is difficult to determine whether a dietary item can affect a clinical problem such as the development of cancer.

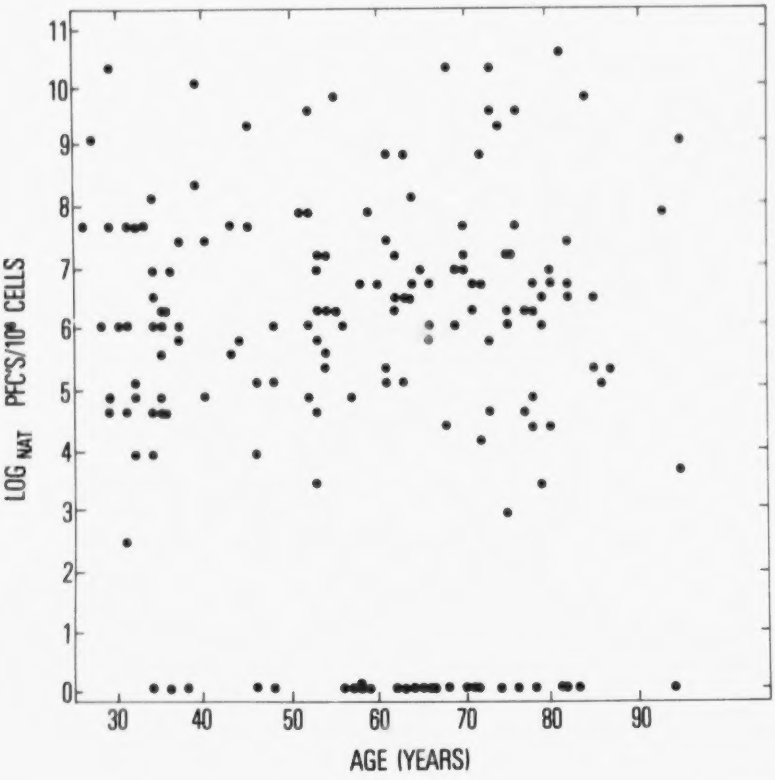


Figure V.34. In these experiments the number of immunoglobulin-producing cells (PFCs) induced by Pokeweed Mitogen stimulation of lymphocytes from BLSA participants of varying age was determined. Each point represents the mean PFC number for replicate PWM-stimulated cultures from each individual.
From Adler and Nagel (1981).

164

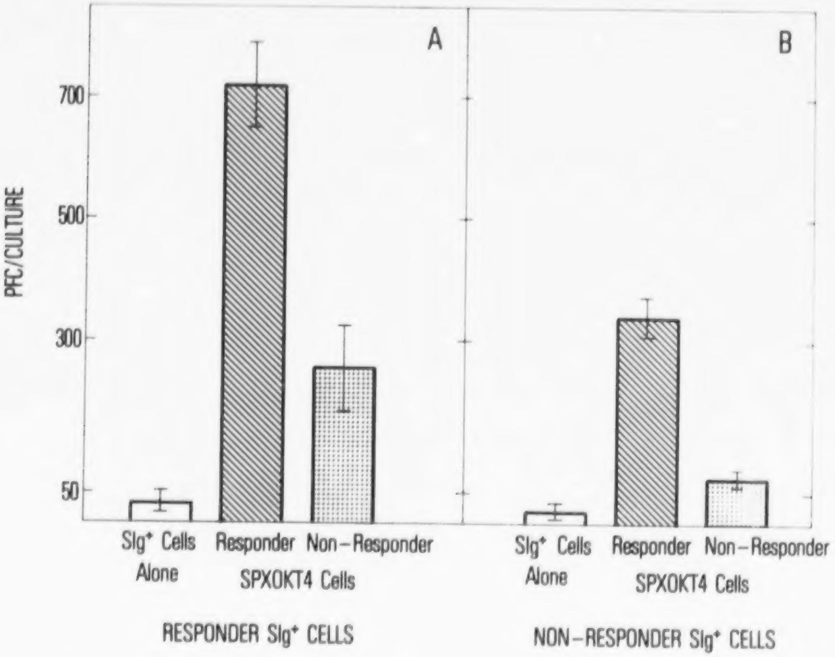


Figure V.35. B cells from a responder (panel A) or a nonresponder (panel B) were cultured alone, or with active helper T cells from a responder or a nonresponder individual. The bars represent the mean PFC number in replicate cultures ± 1 S.D.
From Chrest et al. (1983).

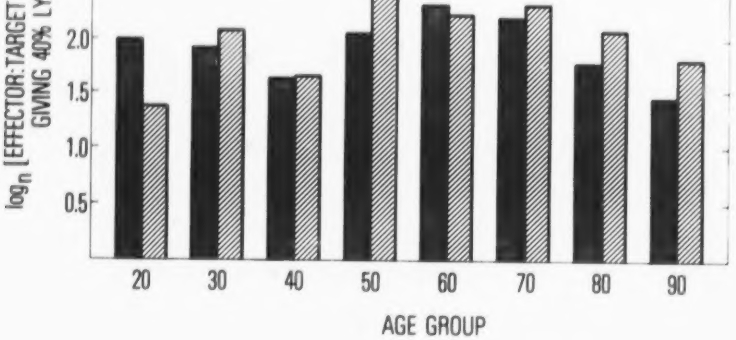


Figure V.36. The geometric mean percentage of the ^{51}Cr release at a 5:1 effector:target ratio was determined as a function of age and sex. There were no statistical differences between sexes or among ages.
From Nagel et al. (1981c).

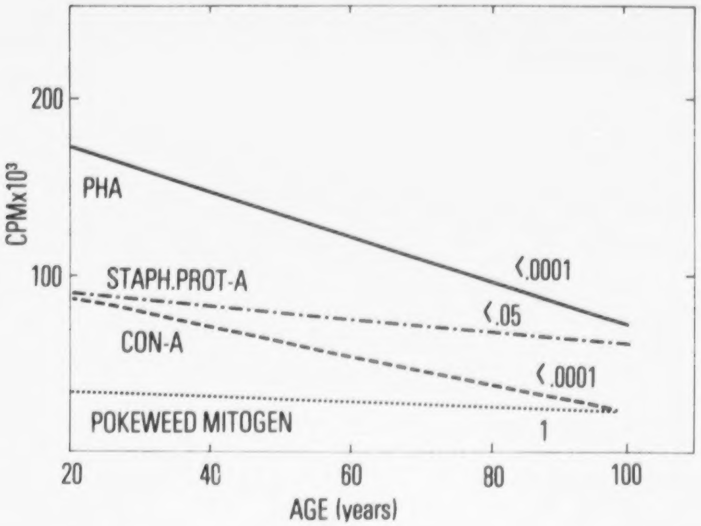


Figure V.37. PBL were stimulated *in vitro* with phytohemagglutinin (PHA), Concanavalin A (Con A), Pokeweed Mitogen (PWM), or Staphylococcal Protein A (SPA). After 5 days of culture the cells were pulsed with tritiated thymidine and one day later the level of radioactivity in the cellular DNA was determined. The P values for each regression line were: PHA = < .001, Con A = < .001, SPA = < .05, and PWM = 1. The regression lines were drawn using the mean values of replicate cell cultures obtained from blood samples from 230 individuals.
From Adler and Nagel (1981).

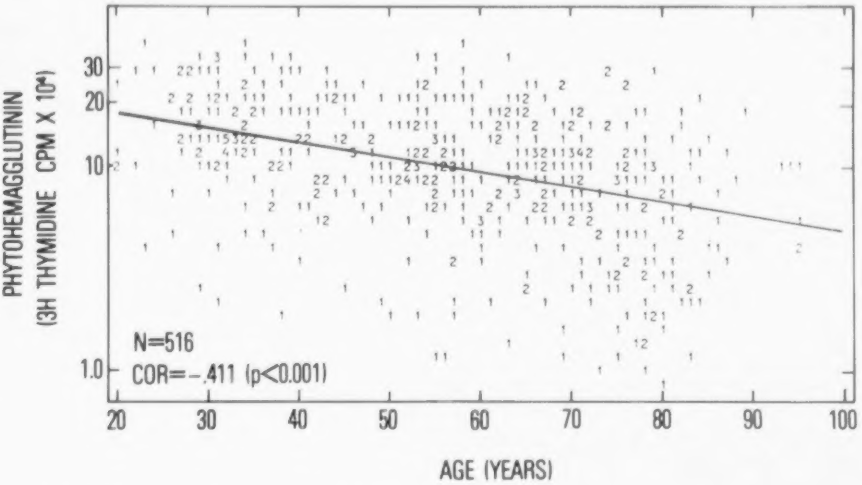


Figure V.38. These experiments are exactly as described in Fig. 37, with PHA used as the mitogen. The data are presented as the mean cpm for the peak response of cell cultures from individuals of varying age. The numbers at each data point show how many individuals of that age had the same amount of thymidine incorporation in their cell cultures.

Mitogen responses of peripheral blood cells. The ability of some PBL to divide *in vitro* in reponse to a mitogenic stimulus makes it possible to determine the presence and functional activity of these cells. Different types of PBL respond in varying degrees to different mitogens. In general, the mitogens phytohemagglutinin (PHA) and concanavalin A (Con A) stimulate T lymphocytes, while Staphylococcal Protein A and PWM stimulate both T and some B lymphocytes. A study of the age of the cell donor and the responsiveness of their cells *in vitro* to a mitogenic stimulus showed that the ability of the T lymphocyte to divide was significantly reduced as the age of the cell donor increased (Adler and Nagel, 1981; Nagel et al., 1982a) (Fig. 37). Some cell cultures from the youngest BLSA age group respond poorly to a PHA stimulus, while some from the oldest age group respond well (Fig. 38). A difference has, however, been found among age groups. The number of poorly responding cultures increases and the number of good reponses decreases with age. The age at which the number of high- and low-responder cultures seems to change markedly is the seventh decade. The populations below age 60 are similar to one another, as are those over age 70 (Fig. 39). Although the reason why the T cells appear to lose the ability to divide *in vitro* is still obscure, it is not a decline in the numbers of T cells. An inherent defect in the T cell must limit its proliferative ability.

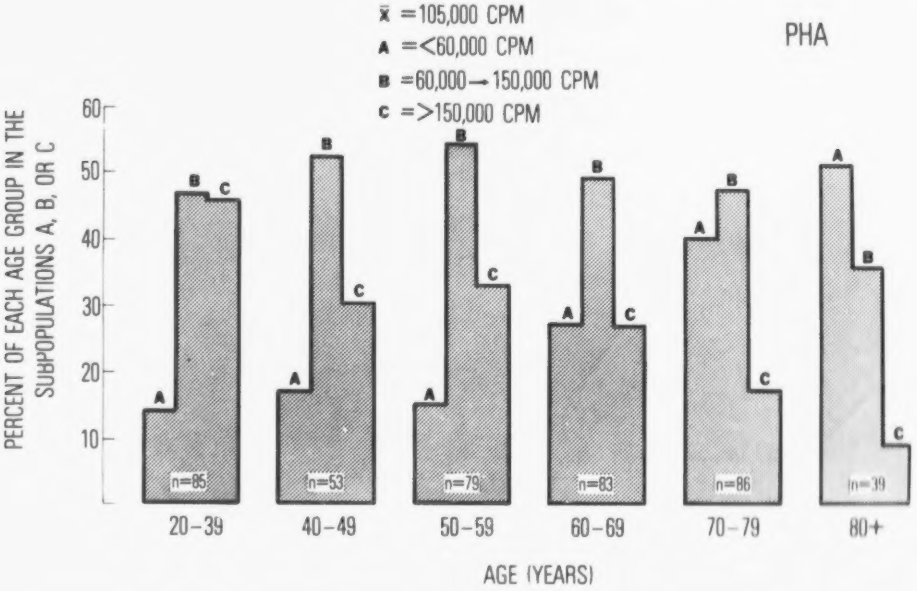


Figure V.39. The experiments are as described in Fig. 37 and in the separation of the data presented in Fig. 38. For each decade of age the individual culture results were divided into 3 groups. One group was composed of results with less than 60,000 cpms (Group A); Group B had results from 60,000 to 150,000 cpms; and Group C had results over 150,000 cpms. The mean cpm for all the data was 105,000.

Serum immunoprotein levels. Immunoglobulins comprise several subclasses: IgG, which constitutes most of the serum immunoprotein; IgM, a macroglobulin immunoprotein; and IgA, a serum immunoprotein that can also be found in secretions. When these were quantified across the age range, no significant differences with age were seen in their representation in serum, except possibly in the group 90+ years old (Adler and Nagel, 1981). The group has so few members, however, that the findings are heavily influenced by a few individuals with serum immunoprotein levels much different from the "normal" levels (Fig. 40).

Summary. The most consistent age-related immunologic findings are the defects seen in the T lymphocyte. The representation of T cells in peripheral blood decreases with age. Both the functional ability of T helper cells and the proliferative ability of T cells in response to a mitogenic stimulus decline with age. It is interesting to speculate that these changes are related to the age-associated involution of the thymus seen in humans and animals.

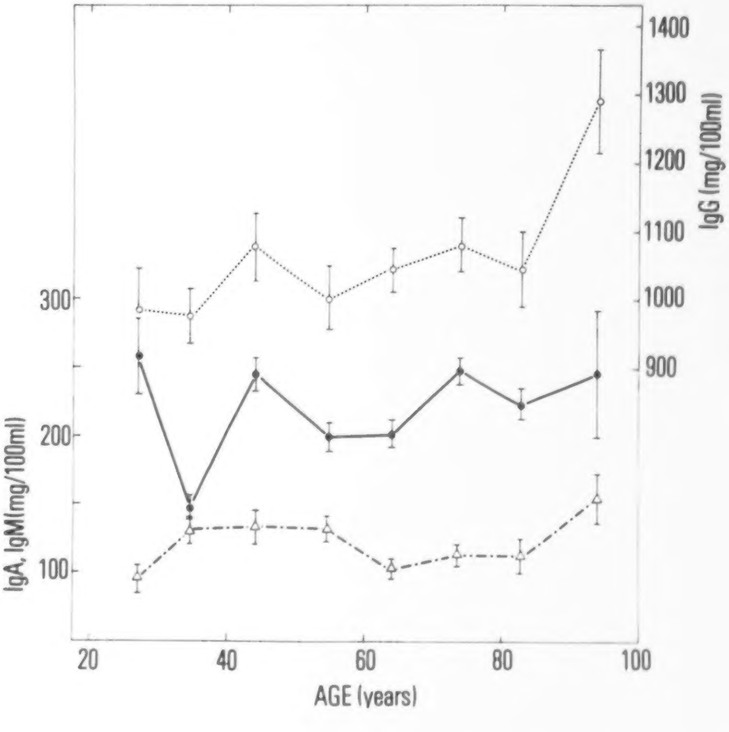


Figure V.40. Serum immunoglobulins G, A, and M were determined on samples from 250 individuals of varying age. The left ordinate refers to the IgA and IgM values, the right ordinate to the IgG values. The data points and standard error bars represent the mean value for a 5-yr interval and are plotted in the mean position for that age group. ○···○ = IgG; ○—○ = IgA; △·····△ = IgM. From Adler and Nagel (1981).

10. Oral/Dental

General dental characteristics. The participants in the oral physiology component are a substantially dentate group, even in the upper age category. The total number of natural teeth in persons aged 60+ is about 23 of 28 maximum (Baum, 1981a). These persons have regular dental care; about 80% visit a dentist at least once a year for preventive as well as problem-oriented care. The prevalence of cervical caries is markedly elevated in older persons; those 40 and younger average one affected tooth in 28, those 60+ four in 28. The average gingival-disease index was quite similar over age groups, while indices of periodontal disease showed a significant increase with age.

Salivary-gland function. Detailed evaluation of stimulated parotid-gland secretion found no difference in the ability of non-medicated men and women of various ages to secrete parotid fluid after stimulation by 2% citric acid (Baum, 1981b). Among subjects taking prescription medication, post-menopausal women showed significantly lower stimulated parotid flow than their non-medicated counterparts.

The secretion of acinar-cell exocrine proteins was studied by following the concentration, output, and percentage in saliva of the anionic-proline-rich proteins. These molecules have been localized to human parotid acinar cell secretory granules. No differences in any of these parameters were observed among non-medicated males and females of any age category (Baum et al., 1982).

The secretion of several inorganic ions (Na^+, Ca^{++}, K^+, and inorganic phosphate) was also examined. Of these electrolytes, only Na^+ showed a consistent alteration with age. Non-medicated men and women both showed significantly lower levels of secreted Na^+ with age. In men 60 and older, the average secretion of Na^+ was about 50% of that of men 40 and younger. These differences suggest increased Na^+ reabsorption by gland ductal cells in older persons.

Gustatory function. Detection thresholds for each of the four basic taste qualities (sweet, sour, salty, and bitter) were obtained from 81 men and women of varying ages. Modest quality-specific age differences were observed. Sodium chloride (salty) showed a higher detection threshold with advancing age ($r = 0.38$), as did quinine sulfate (bitter), although its magnitude ($r = 0.25$) remained lower than that of salt. No age differences were observed for either citric acid (sour) or sucrose (sweet).

Motor function. Physical diagnostic evaluations observed specific decrements in certain oral motor functions with age. The frequency of altered masticatory function was significantly greater with age among both men ($X^2 = 16.3$) and women ($X^2 = 23.2$). The postural functions of the circumoral muscles appeared similarily diminished with age (men, $X^2 = 9.51$; women, $X^2 = 4.52$), while postural function of the tongue musculature showed a higher frequency of age-related dysfunction among men ($X^2 = 12.4$) but not among women (Baum and Bodner, 1983).

BEHAVIOR

1. Psychophysiology

Effect of Age on Reaction Time and Vigilance

Cross-sectional studies have shown that aging is associated with a reduction in the speed of responses and an increase in the probability that the response will be inaccurate. In short, with aging speed diminishes and errors increase.

A number of studies also showed that the effects of age became greater as the complexity of the responses increased. Tasks that required the subject to choose among a number of responses showed greater impairment (more time required with more errors) than tasks eliciting a single response.

In some instances physiological indices could be included as a measure of the response. Thus reaction time, latency of response, vigilance, EEG response, heart rate, blood pressure, galvanic skin response (GSR), and skin potential were examined in subjects of different ages in a search for relations between behavioral and physiological events.

A study was undertaken to test the effect of age on reaction time (Suci et al., 1960). Old (age range = 60–70 yr, median = 63) and young (age range = 17–38 yr, median = 18.5) subjects were individually tested for choice reaction time by an apparatus consisting of four small electric lights mounted on a panel. A varying number of lights was illuminated; when one was turned off, the subject was asked to name it, using a previously assigned code name, as quickly as he could. A plot of reaction time against bits per stimulus in the two age groups showed that reaction time in both old and young subjects increases linearly with the number of stimuli from which the subject must choose, and that age differences in reaction time increase as a function of increasing stimulus information; older subjects show longer reaction times at all levels of stimulus information (Fig. 41).

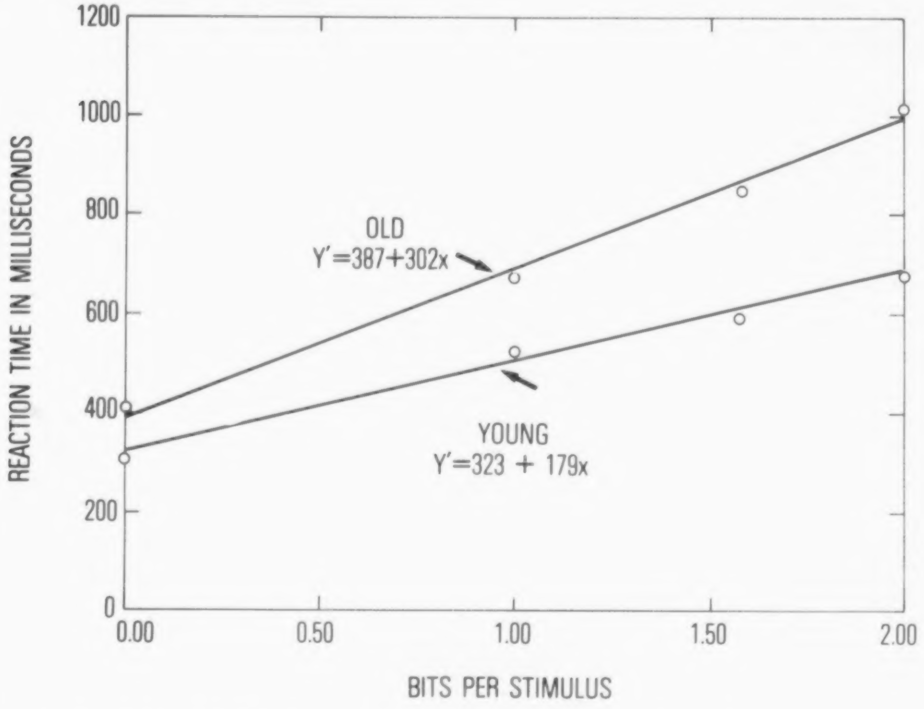


Figure V.41. Reaction time as a function of stimulus information and age. From Suci (1960).

A study was undertaken to determine the relation of vigilance to age and whether lowered vigilance is associated with age-related slowing in reaction time (Surwillo and Quilter, 1964). The subjects, 106 men aged 22 to 82 years, were given Mackworth's Clock-Test. The "clock" is a metal box with a plain circular white face 12 inches in diameter, and a single black pointer six inches long, mounted from the center of the clock. The pointer moves in discrete steps like the second hand of an escapement clock. The full circle is completed in 100 steps, one per second. At long and irregular intervals, the pointer jumps through twice the usual distance in one step. Twenty-three "double jumps" (0.64% of all pointer movements) occur per hour.

After instruction, demonstration, and practice sessions, the subjects were asked to watch the clock-pointer for one hour and to press the response key as fast as possible whenever a double jump occurred. The mean percentage of double jumps detected by the younger group (< 60 yr) was 72.9%, by the older group (> 60 yr) 64.4%. It was concluded that, under the conditions of the experiment, old people were less vigilant than young people. Fractionation of the data into quarter-hour periods showed that the old subjects were not significantly less vigilant than the young subjects during the first 15 minutes of the test, but the mean difference between the first and fourth quarter-

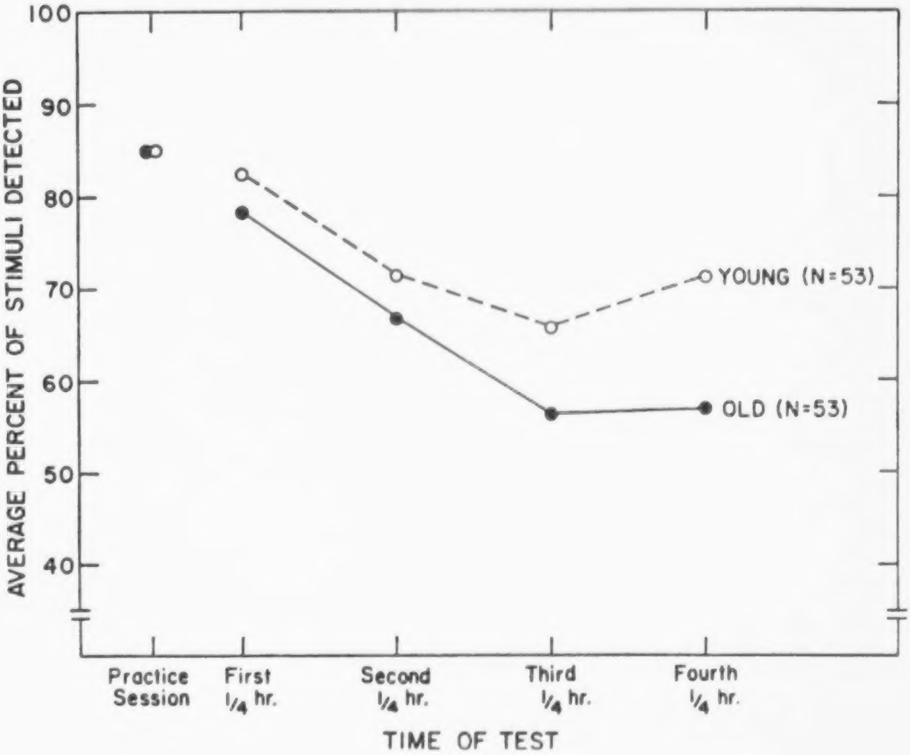


Figure V.42. Decline of vigilance during prolonged visual search. From Surwillo and Quilter (1964).

hour periods was significantly greater in the old group (21.3%) than in the young group (11.2%) (Fig. 42). The conclusion was drawn that vigilance declines more in old than in young persons. At the same time, the lack of an age difference in vigilance during the early periods of the task has important implications for the employment of older men. The effect of short rest periods on vigilance in older subjects deserves further investigation.

Electroencephalographic (EEG) Studies

Several studies have been made of the relation between reaction time and brain waves recorded by electroencephalography.

Since the early 1930s several investigators have advanced the hypothesis that the EEG frequency, which is not constant in an individual but varies from one instant to the next, represents the basic unit of time in the central nervous system. Further testing of the hypothesis was undertaken by Surwillo (1961, 1963a). In a sample of 100 males aged 28 to 99 years ($\bar{x}$ age = 55.3 yr), reaction time to a sound stimulus was related to the alpha frequency at the time the stimulus was given. Stimuli, responses, and EEG were recorded simultaneously. A correlation of 0.72 was obtained between average reaction time and the average period of the EEG (Fig. 43). This coefficient was scarcely altered by exclusion of age from the relation by means of partial correlation. Age was

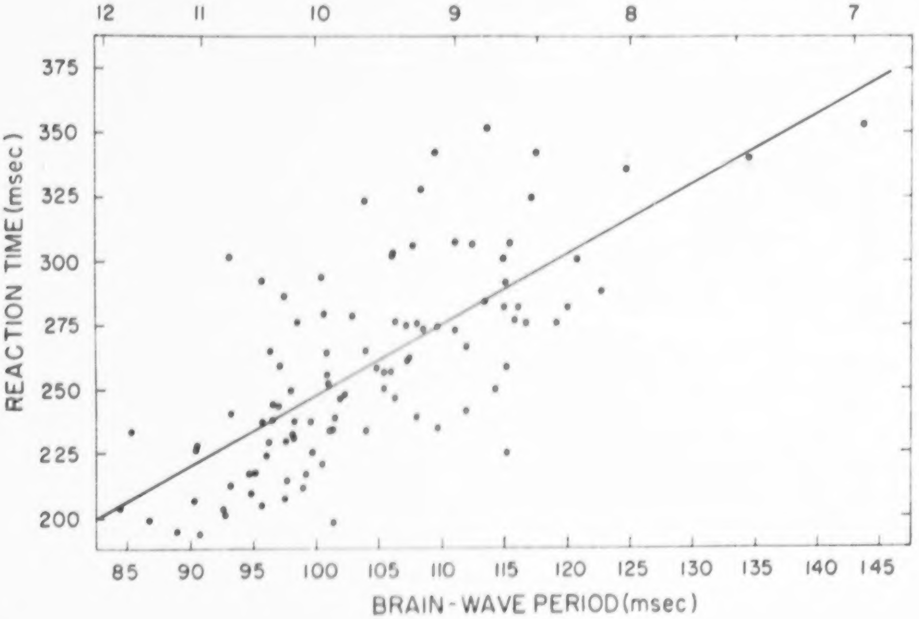


Figure V.43. Reaction time plotted against brain-wave period for data from the high-vigilance, high-motivation condition. Each point represents the average reaction time and average brain-wave period of a single subject, and is derived from a mean number of 14 observations. Numbers at the top of the graph refer to the corresponding frequencies in c/sec. The coefficient of correlation, r, = 0.72; N = 100.
From Surwillo (1963a).

thus not a factor in the observed correlation. A significant positive correlation (r = 0.57) was found between age and the average period of brain waves (Fig. 44). A low but statistically significant positive correlation between average reaction time and age vanished and became negative when brain-wave period was partialled out; the implication is that EEG frequency and not age is the central-nervous-system factor that determines the age-associated slowing in response time. A positive correlation was found in individual subjects between brain-wave period and reaction time. These data are consistent with the hypothesis that the brain-wave cycle is the basic unit of time by which a response is programmed by the central nervous system.

The evidence that variability in response time within subjects increases with age led to a study to determine whether this could be accounted for by differences in variability of brain-wave period (Surwillo, 1963b). A sample of 100 male subjects aged 28 to 99 years ($\bar{x}$ age = 55.3 yr) was tested essentially as in the preceding study. The

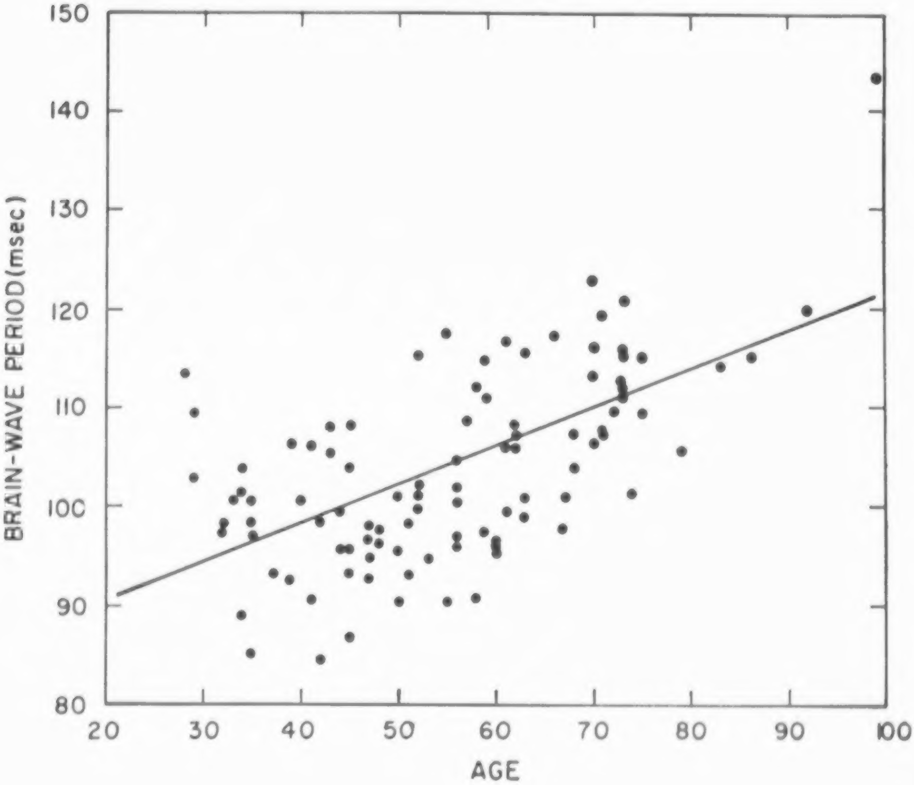


Figure V.44. Period of the electroencephalogram plotted against age. Each point represents the age and average brain-wave period, recorded in the interval between stimulus and response, of a single subject. The correlation coefficient, r, = 0.57; N = 100. Regression line is defined by the expression: P = 0.388 AGE + 82.79, where P = brain-wave period. Standard error of estimate, δ (est. P) = 8.37.
From Surwillo (1963a, adapted).

previous results were confirmed by the finding of a statistically significant positive correlation between age and individual variability in reaction time. Although the relation could not be accounted for by differences in variability of brain-wave period within individual subjects, when average brain-wave period was held constant through partial correlation the positive coefficient relating reaction-time variability and age vanished. The inference was thus drawn that differences in variability of response time are a consequence of differences in average brain-wave period.

The finding that EEG frequency and reaction time are correlated permits no conclusions as to cause and effect. An attempt was therefore made to determine causality by altering the EEG frequency experimentally and measuring the effect on response time (Surwillo, 1964a). It has been observed that repetitive flashing of a light (photic "driving") can produce frequency changes in the EEG; it was assumed for this study that the photic response and the normal brain rhythm are produced by similar mechanisms.

A sample of 48 healthy males aged 34 to 101 years was photically stimulated, and EEGs were recorded as has been described. Subjects whose tracing showed clear evidence of synchronization were given a reaction task to perform while the light was flashing: They were instructed that whenever an auditory signal, which occurred without warning, was presented, they were to press a response key as quickly as possible. Approximately 30 signals were given at random in about ten minutes, while flash frequency was varied. Synchronization of the EEG and the flashing light occurred in 18 subjects, but of these six failed to maintain synchronization while performing the reaction task, six others showed it only over a very narrow frequency range, and one revealed it in only one brain hemisphere; only five of the 18 subjects could synchronize their EEGs with a flashing light over a wide frequency range while performing the reaction task. The findings indicate that at least in some subjects reaction time can be influenced by EEG frequency.

All the EEG studies described above concerned voluntary responses. The generalization that the period of the alpha rhythm is an important factor in the temporal organization of behavior requires that the relation hold for involuntary responses as well. Since studies by earlier investigators had demonstrated latency of attenuation (delay in "blocking") of the alpha rhythm in children and had found a relation between response speed and latency of alpha attenuation, a study (Surwillo, 1966a) was undertaken to determine whether the latency of alpha attenuation increases with increase in alpha-rhythm period independently of differences associated with age. A second purpose of the study was to test the hypothesis that latency of alpha attenuation increases in old age. A final purpose was to test the hypothesis that EEG reactivity, or the incidence of alpha attenuation, declines in old age. Ninety healthy males aged 17 to 91 years ($\bar{x}$ age = 49.7 yr) were tested. The EEG was attenuated by high-intensity flashes of white light lasting approximately one second; 25 flashes in about eight minutes were distributed more or less randomly but occurred mainly during periods when well-defined EEGs were being recorded. Average latency of alpha attenuation and average period of the EEG, in the interval between flash and initiation of the involuntary response, were determined along with the number of times the stimulus failed to attenuate the EEG. A significant positive relation, independent of age, was found between attenuation latency and EEG period. A low but statistically significant positive correlation was found between age and latent time of alpha attenuation. A low correlation was also found between EEG reactivity and age.

Although it was concluded that age cannot account for the relation between attenuation latency and EEG period, the finding confirms earlier work suggesting that old persons are as a group less activated or aroused than young persons by sensory stimulation. It remains to be determined whether this represents an age difference at the level of the cerebral cortex.

Electrodermal Activity

Electrodermal activity (EDA) includes both skin conductance and skin-potential phenomena. BLSA studies have focused an age differences in skin-potential response (SPR) and level (SPL), and skin-conductance response (SCR) and level (SCL). Although age differences in EDA may reflect age differences in responsiveness of the autonomic nervous system, they also may reflect age differences in the peripheral sudomotor system.

Skin-potential response and vigilance. The frequency of spontaneous SPRs in relation to vigilance and age was examined in a BLSA study (Surwillo and Quilter, 1965b) of 132 healthy males, aged 22 to 85 years. The subjects were given Mackworth's Clock-Test described above ("Effect of Age on Reaction Time and Vigilance"). Their task was to press a response key as quickly as possible whenever a double jump occurred. If the subject did not press the response key within six seconds of the appearance of a double jump, it was assumed that he had not perceived the signal. Skin potential between the palm and the ventral surface of the left forearm was recorded from Ag-AgCl electrodes. The primary datum in the investigation was the number of SPRs (negative deflections only) that occurred within the 18 seconds immediately preceding the double jump. This frequency was considered a measure of vigilance, since detected double jumps were preceded by significantly more SPRs ($\bar{x} = 1.62$) in this period than were undetected double jumps ($\bar{x} = 1.22$).

The frequency of spontaneous SPRs was lower with age. Mean values for 42 subjects aged 22–47, 47 subjects aged 48–67, and 43 subjects aged 68–85 years were respectively 1.70, 1.52, and 1.30 SPRs.

Effect of age on the latency of involuntary and voluntary responses. Before 1965, numerous investigations of voluntary reaction time showed a general increase in response latency with increasing age. The latency of involuntary responses, on the other hand, had not received much attention, nor were the findings clearcut or consistent. A BLSA study was therefore undertaken to measure simultaneously the latency of voluntary and involuntary responses to the same stimulus (Surwillo and Quilter, 1965a). Involuntary response latency was measured by the latency of the galvanic skin response (GSR), while voluntary response latency was the reaction time. The aim was to see whether old persons show greater latencies than young persons on both measures.

One hundred thirty-two healthy males, aged 22 to 85 years, were classified in three age groups: Group I (42 men aged 22–47 yr); Group II (47 men aged 48–67 yr); and Group III (43 men aged 68–85 yr). As in the research described in the preceding section, the subjects performed an hour-long watchkeeping task in which they monitored the movements of a clock pointer and responded to double jumps by pressing a key. GSRs (negative deflections only) and the instant of a subject's voluntary response were recorded on the same chart. Lack of an involuntary or voluntary response within six seconds of a double jump was considered an indication of failure to perceive the stimulus.

The mean GSR latency for Group I was 1.81 seconds, that for Group II, 1.82 seconds; the latency for Group III, 1.95 seconds, was significantly higher. The latency of voluntary responses to the same stimuli, however, showed no increase with advancing age.

Effect of age on skin-potential level (SPL). A BLSA investigation was made to determine whether age differences also occurred in SPLs (Surwillo, 1965). One hundred twenty-two healthy males, aged 22 to 85 years, were selected for SPL measurements. SPLs between the palm and ventral surface of the forearm were recorded from Ag-AgCl electrodes during the first 15 minutes of the previously described hour-long vigilance task. (Although previous studies had found old persons less vigilant in performance of the entire task, age differences were not apparent during the first 15 minutes.) SPL was measured at 42 different points in the 15-minute recording, and the mean of these measures was calculated.

In all subjects, the palm was always electrically negative with respect to the forearm. Individual values of skin potential ranged from -12.3 to -56.8 mV (average for the group was -31.5 mV). A low but statistically significant negative correlation was found between age and SPL.

Relation of autonomic activity to age differences in vigilance. Because autonomic activity is related to the level of activation, a BLSA study was performed to determine whether differences in autonomic activity are associated with the age differences in vigilance (Surwillo, 1966b). The hypothesis tested was that, during the vigil (the Mackworth's Clock-Test), measures of autonomic activity (heart rate, palmar skin temperature, and palmar skin potential) change at different rates in old and in young persons.

Sixty-six healthy males were divided into two age groups of equal size, 22 to 45 years ($\bar{x}$ age = 36.4 yr) and 69 to 85 years ($\bar{x}$ age = 74.3 yr). The subjects' heart rates, palmar skin temperatures, and SPLs were recorded during the 18 seconds preceding each double jump. In the final 45 minutes of the task, heart rate declined and SPL became more negative. Skin temperature, however, declined progressively in the young but not in the old group, a finding that is consistent with the hypothesis that differences in autonomic activity are associated with the more rapid decline of vigilance in old subjects.

Effect of epidermal hydration on skin potential and conductance. Although age differences in EDA may reflect age differences in activation of the sympathetic nervous system, age differences in the peripheral sudomotor system also influence the results. Edelberg's (1968) model illustrates how peripheral characteristics can influence the expression of nervous activity. Of critical importance is the ratio of epidermal resistance to sweat-gland resistance. The larger the ratio the greater the sweat-gland contribution to the electrodermal recording. Hydrating the epidermis lowers the epidermal/sweat-gland resistance ratio and thus reduces the sweat-gland contribution to EDA (Edelberg, 1968).

The following studies demonstrate that age differences in SPL occur only when a nonhydrated recording site is used (i.e., when conditions maximize the sweat-gland component). The finding may reflect a less negative sweat-gland potential in old age and/or age differences in epidermal/sweat-gland resistance ratio.

Garwood et al. (1979) studied 12 young men (age range = 23–36 yr, $\bar{x}$ age = 30.75 yr) and 12 old men (age range = 63–82 yr, $\bar{x}$ age = 75.50 yr). Three skin-hydration conditions were produced on sites used for electrodermal measurements. In order of increasing hydration, they were: a) 0.5% KCl-glycol electrolyte, b) 0.5%

KCl-agar electrolyte, and c) presoaking with distilled water followed by 0.5% KCl-agar electrolyte.

Three tasks were given to the subjects. In task 1, the subjects received eight presentations of a 50-dB, 250-Hz tone. In task 2, a reaction-time task, the subjects received a 50-dB, 250-Hz warning signal, then a 50-dB, 1000-Hz signal, to which they had to respond by pressing a key. Task 3 was a choice-reaction-time task in which the warning signal was again 50-dB, 250-Hz, which was followed either by a response signal of the same loudness and frequency or by a 50-dB, 1000-Hz tone; the subjects were to respond only when the low-pitch warning signal was followed by another low-pitch tone. During both reaction-time tasks the subjects were repeatedly urged to respond more rapidly.

There were no significant age differences in the effect of electrolyte medium on SCL and SCR. The older subjects had lower SCL and SCR magnitudes than the younger subjects. There were age differences in the effect of electrolyte medium on SPL and SPR. The SPR in young adults was related to hydration, the largest response occurring with the least hydration, but electrolyte did not significantly affect the magnitude of SPR in the old subjects. Age differences in SPL and SPR occurred only with the glycol medium, the older subjects having smaller responses and less negative levels. The authors postulate that the reversal in hydration/SPL relation with age reflects a reversal in the relative magnitudes of the potentials of epidermis and sweat glands: In old persons the epidermal is greater than the sweat-gland potential.

Garwood et al. (1981), reporting that differences in SPL were dependent on epidermal hydration, also found that age differences occurred only when SPL was more negative than resting levels. Basal skin-potential-level (BSPL) procedures were used. Subjects were required to relax for 30 minutes until a minimum SPL (BSPL) was reached during an absence of electrodermal response. BSPL is considered to be a nonsudorific SPL component (Christie and Venables, 1971). Hydration conditions were as in Garwood et al. (1979). Again, age differences occurred only with the glycol medium. Furthermore, age differences occurred only at the start of the recording, when sweat-gland activity can be expected to contribute to SPL. There were no significant age differences at BSPL. Capriotti et al. (1981) also reported that with increasing age there is a decline in SPL negativity when the glycol medium is used.

An additional consideration in investigating age differences in EDA is recording site. Capriotti et al. (1981) reported that age differences in SPL were much larger when recordings were made from the medial phalanges than when they were made from the thenar eminence.

Time Perception and Psychomotor Function

Effect of age and institutionalization on perception of short intervals of time. While the common observation that time seems to pass more quickly as one ages has received some experimental support, the hypothesis has been challenged by some investigators. An investigation was undertaken to clarify the issue experimentally (Surwillo, 1964b). The study included a comparison of estimates of short intervals of time made by non-institutionalized subjects with those of an age-matched group of institutionalized subjects.

One hundred twenty healthy white BLSA males 27 to 90 years old were classified in three age groups of 40 subjects each with means of 37.5, 56.1, and 73.7 years. A group of 40 white males from the Infirmary Division at BCH with a mean age of 73.7 years (± 6.2) was compared with the old group of healthy subjects. The institutionalized subjects were low-income patients who were active, could feed, wash, and dress themselves, and had no apparent psychiatric disorder.

Each subject was asked to estimate intervals of 30, 60, and 180 seconds by holding a telegraph key closed for the stated period of time. Elapsed time was measured with a chronoscope that was activated by the telegraph key. Three estimates obtained for each interval were averaged to determine the subject's score.

Among the non-institutionalized subjects, there was no systematic increase with age in estimates of time intervals, and none of the mean estimates differed significantly from actual elapsed time. The hypothesis that a short interval of time is perceived by older persons to move at a faster rate was therefore rejected. The fact that the time estimates made by the group of institutionalized subjects were significantly shorter than those of age-matched non-institutionalized subjects may possibly reflect differences in the "content" of time for infirmary patients; it may equally well reflect differences between the two groups in education, intelligence, and socioeconomic status.

Speed and accuracy of movement. The effect of age on speed and accuracy of movement was studied (Welford et al., 1969) in subjects who tapped a pencil back and forth as quickly as possible between two targets each consisting of two parallel lines drawn on paper. The target widths (the distances between the parallel lines) were 32 mm, 11 mm, and 4 mm. The distances between targets—from the center of one to the far edge of the other—were 50, 142, and 402 mm for each width. Observations of each subject were fitted to equations, based on information theory, that relate movement time to average distance from starting to finishing point and an estimate of the scatter of the points around each target (a measure of accuracy).

Data from 325 men ranging in age from 20 to 70 were analyzed in ten-year age groups. Performance, as measured by the time required to place 50 dots on each target, improved from age 20 to age 40 and then declined. At any age more time was required to place the dots in the narrow than in the wide targets. The greatest accuracy in positioning was achieved by the 20-year-olds. Analysis of the data led to the conclusion that two control processes could be distinguished—one controlling movement over distance (motor control) and a slower one that controlled "homing" onto the target (visual control). The authors concluded that position sense and motor control decline more with age than do the decisional processes responsible for homing on a target.

Muscle strength and coordination. Arm strength and manual cranking ability were measured in 218 BLSA participants 20 to 89 years of age (Shock and Norris, 1970). The maximum power output generated in the crank-turning task, which required coordination of several muscles, was compared with the strength of the same muscles when their movement was kept to a minimum (i.e., isometric strength measurement). In contrast with previous studies in large unselected populations, in which a decline in muscle strength was evaluated by strength of grip (Fisher and Birren, 1947; Welford, 1959, 1977), the study found that the strength of arm and back muscles in these selected subjects did not differ significantly over the age span from 20 to 65 years (Fig. 45). Significantly lower muscle strength was found only in subjects in their 70s and 80s. The 28% difference in muscle strength found in the 80-year-olds was reduced to 19% when allowance was made for the lower body weights of this age group. While muscle strength showed no significant differences until the last two decades, power

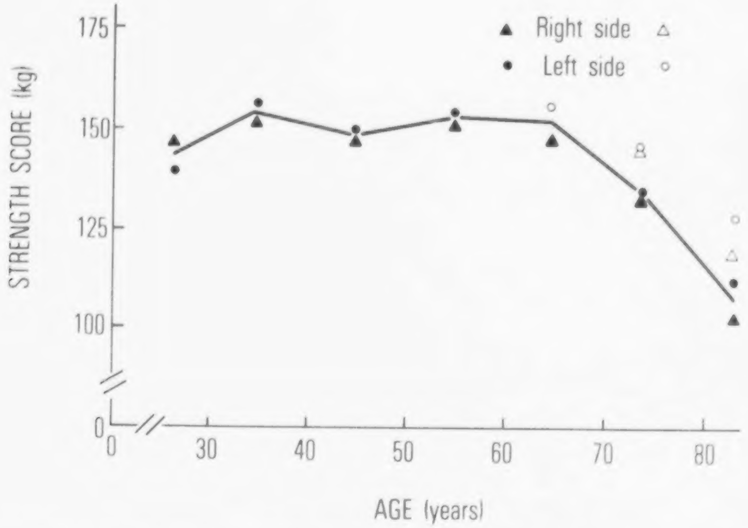


Figure V.45. A composite strength score (kg) for the arm and shoulder muscles is plotted against age (yr) for the right side (▲) and the left side (●). Values that would be expected if the oldest subjects had weighed as much as younger subjects are shown for right side (△) and left side (○).
From Shock and Norris (1970).

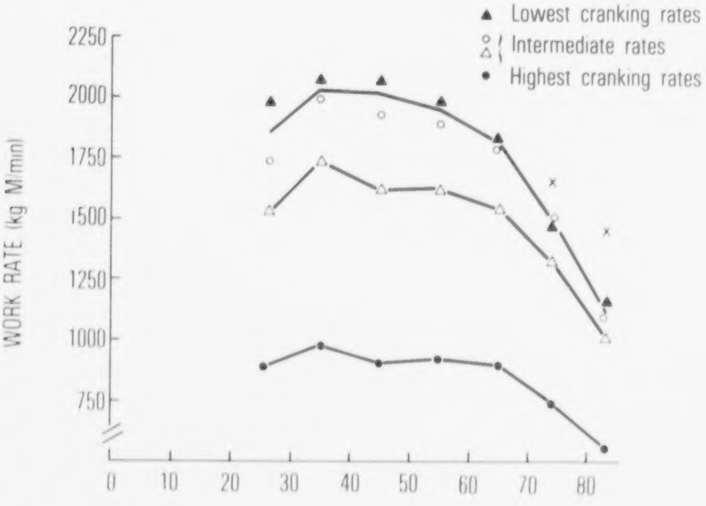


Figure V.46. Maximum work rate (power output) (kgM/min) is shown for 4 cranking rates for each of 7 age-decade groups. Values that would be expected if the oldest subjects had weighed as much as the younger subjects are shown for the 2 lowest cranking rates (X). From Shock and Norris (1970).

output showed significant differences as early as the fifth decade. The largest difference, 45% among 80-year-olds, was reduced to 30% when allowance was made for lower body weight (Fig. 46). Since the older subjects were highly motivated to achieve maximal results and the duration of the cranking exercise was too brief to manifest any cardiovascular or pulmonary limitations, the most likely cause of their power-output deficit was judged to be reduced coordination ability.

2. Cognitive Functions

Learning and Memory

Age differences in verbal learning. Although several studies have indicated that performance in verbal learning declines with age, it was not until the 1960s that the rate of presentation, or pace, of verbal information was evaluated as an independent variable. Since some errors in paced trials may be due to insufficient time to respond, rather than to failure to learn, direct resolution of the learning-performance problem requires measures of errors that are not attributable to pace. To provide such measures, a study of age differences in paired-associate learning (Arenberg, 1965) included self-paced trials (without feedback) alternated with paced trials. In addition, the anticipation interval (the time allowed for response) was independent of the inspection interval. In this way, differences between pace groups would not be attributable to variations in the time allowed to view the material to be learned. It was hypothesized that the old subjects would make more errors than the young, that age difference would be greater at the faster than at the slower pace, and that the pattern of self-paced errors would be similar to the pattern of paced errors despite the fact that each subject could take as much time as he needed to respond.

In Study I the subjects were 64 men participating in the BLSA. Participants were randomly assigned to the fast or slow pace; 32 men were 29–40 and 32 were 63–77 years old. On each trial the eight items were presented in one of five orders. The stimulus component of each item consisted of two consonants; the response component was a familiar two-syllable adjective. The material was presented by means of a card-changing instrument equipped with variable control of the shutters. The anticipation interval, during which the two consonants were exposed, was 1.9 seconds for the fast pace and 3.7 seconds for the slow pace. For both pace groups the inspection interval, during which consonants were exposed together with the word, was 1.9 seconds. The interval between the displays of items was 1.8 seconds for both groups. The procedure continued until one errorless trial or 52 trials occurred.

Although the mean number of errors during the fast-pace test was substantially greater in old than in young subjects (Fig. 47), it was impossible to judge whether the older subjects had difficulty in learning at that pace or had learned the material but could not respond in the short anticipation interval. Study II was designed to determine which explanation was correct.

Subjects in Study II were unemployed men, not members of the BLSA, with at least a sixth-grade education, who were seeking work at a state employment agency. Sixty-four men with raw scores of at least 20 on the WAIS Vocabulary test were recruited. The age ranges were 18 to 21 years for the young group and 60 to 77 years for the old. Subjects were randomly assigned to the fast or slow pace to provide 16 subjects in each of four age-pace groups. The eight-item list was reduced to six items.

To provide measures of errors not attributable to insufficient time to respond, self-paced test trials were alternated with paced acquisition trials for each subject. Timing of the paced trials was the same as in Study I. As with the better-educated subjects of Study I, the older group showed a significantly greater frequency of errors in the paced trials. Older subjects also committed more errors in self-paced trials, an indication that their poorer performance was due to a learning deficit rather than to an inability to respond quickly.

In another study (Arenberg, 1967b), men in the BLSA whose age range was 20 to 87 years were given a paired-associate learning task (335 men) and a serial learning task (322 men). The card-changing instrument and the eight-item list of the previous study were used for the paired-associate learning task; the anticipation, inspection, and between-item intervals were the same as before. The instrument was also used, with different time intervals, for the serial learning task, in which 12 familiar five-letter words were shown to each subject in a fixed order.

Polynomial regression analysis elicited linear components at both long and short anticipation intervals, but quadratic components only at the short interval. The

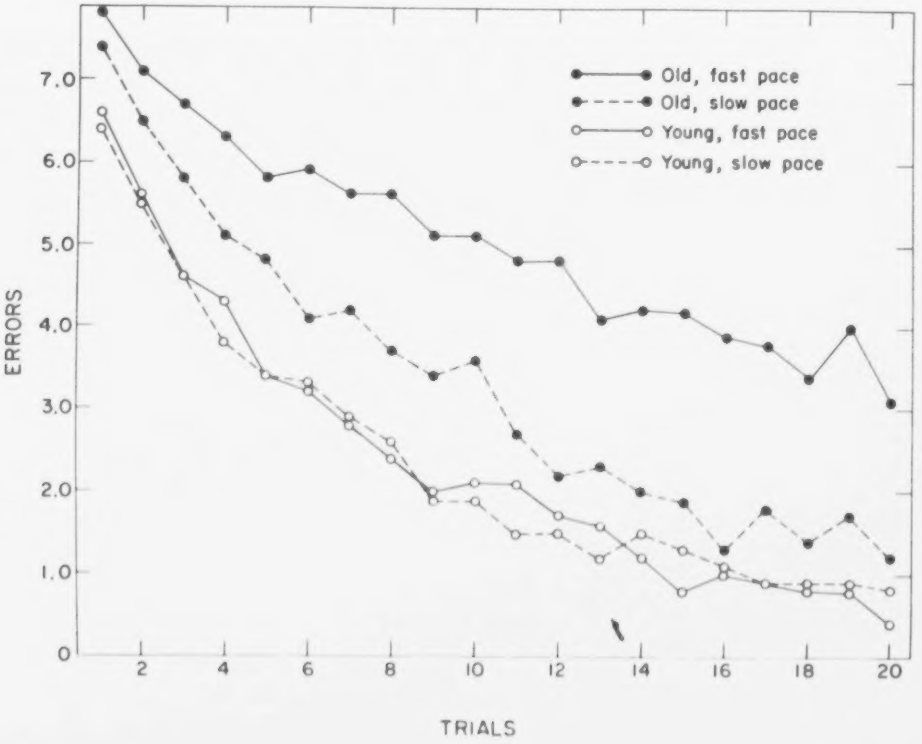


Figure V.47. Mean errors for all 4 age-pace groups plotted over the first 20 trials. From Arenberg (1965).

quadratic component indicated that the magnitude of age differences increased with age. It is possible that each passing year results in a larger decrement than the preceding one in an individual's performance; it is also possible that some threshold level of functioning of the underlying mechanisms must be reached, or some event must occur, in order to produce a performance impairment, and that the impairments are greater as age increases. The issue can be resolved only by longitudinal measurements.

Age differences in retroaction. "Retroaction" refers to the effect of an interpolated task on the recall or relearning of previously learned material. Two previous studies of age and retroaction had produced contradictory results. Gladis and Braun (1958), who used long anticipation intervals, found no age differences in retroaction, while Wimer and Wigdor (1958), who used short intervals, found that old subjects were affected more than young by the interpolated activity. A BLSA study (Arenberg, 1967a) used two anticipation intervals in a paired-associate learning task to test the possibility that the contradictions were due to the use of very different anticipation intervals. It was hypothesized that an age difference in retroaction would be found at the short anticipation interval but that no age difference, or a small one, would be found at the long interval.

The subjects were 24 young (age range = 30–39 yr) and 24 old (age range = 62–77 yr) BLSA participants. The men were randomly assigned to the short (1.9 sec) or long (3.7 sec) anticipation interval. Approximately two minutes elapsed between original learning and interpolated learning and between interpolated learning and relearning. An age difference in relearning was found at the short but not at the long anticipation interval. The results were thus in essential agreement with the earlier studies: Some part of the difference in those findings was apparently due to a difference in the anticipation interval.

Age differences in memory and decision performance. A study was conducted to assess the magnitude and pattern of age differences in experimental laboratory measures of cognitive and psychomotor performance (Robertson-Tchabo and Arenberg, 1976). Performance measures for a wide variety of cognitive laboratory tasks, mostly memory and decision tasks, were factor-analyzed. The sample consisted of 96 healthy BLSA men whose age range was 20 to 80 years. The factors were identified as speed of information processing, secondary memory, attention, and primary processing efficiency. All factor scores were correlated with age, better performance being associated with lower age. Attention was found to have the highest, primary processing the lowest correlation with age. The sample was divided into subsamples of 32 young (20–39 yr), 32 middle-aged (40–59 yr), and 32 old (60–80 yr) subjects, and each subsample was factor-analyzed separately to determine whether the factor structure was similar for all age groups. Evidence was found of factor-structure invariance with adult age; all four factors in the primary analysis were identifiable in each age subsample. The findings are consistent with a model of continual cognitive decline with age in healthy, educated adult males. The decline seems to be quantitative rather than structural in nature: Age appears to function as a scalar affecting the magnitude of a factor score.

Equivalence of Information in Concept Identification

A new method was designed at the GRC to provide equivalent amounts of information for logically equivalent selections in concept identification when subjects select instances (Arenberg, 1970). An "instance" is an item or stimulus configuration in a concept-identification task. A positive instance exemplifies the concept, that is, it includes the defining attributes of the concept; a negative instance does not exemplify the concept. Equivalence is maintained between comparable conjunctive and disjunctive problems as well as within a particular type of problem. The equivalence-of-information method permits the amount of information gained to be used to measure performance for each selection. The non-equivalence of available information, which usually results when subjects select instances, is avoided by this method. Data from an experiment in which the method was used suggest that the amount of initial information provided merits investigation as a variable. Moreover, pooling the results of concept problems with high and low initial information may obscure important group differences.

3. Daydreaming

Age differences in daydreaming. A BLSA study of daydreaming across the adult age span (Giambra, 1974) investigated the frequency, content, temporal setting, and precipitating instances of daydreaming. Daydreams were defined as spontaneous thoughts, unrelated to the task at hand, that intrude into the person's awareness.

Six age groups of BLSA participants were studied: 24–34 (n = 20), 35–44 (n = 13), 45–54 (n = 31), 55–64 (n = 46), 65–74 (n = 28), and 75–91 years (n = 26). In addition, a young group of college students 17 to 23 years old (n = 214) were recruited. The Imaginal Processes Inventory (IPI), a 28-scale questionnaire, was used to measure aspects of daydreaming and related activity. Linear declines with age were found in daydreaming frequency; absorption in daydreaming; imagery and vividness in daydreams; and daydreams about sex, bizarre and improbable events, achievement, hostility, heroism, fear of failure, and guilt. No decline was found in acceptance of daydreams, impersonal and interpersonal curiosity, mentation rate, past and present temporal setting in daydreams, or daydreams involving problem solving. Except in the youngest group, in which sexual daydreams predominated, problem-solving daydreams were predominant at every age. There was no concentration on daydreams about the past in any age group, including the oldest.

The study was later replicated (Giambra, 1977–78). The characteristics of daydreaming obtained in the original sample were also found in the replication, which thus supported the earlier findings. Combination of the original and replication samples allowed a closer analysis and provided tentative norms for the populations sampled. Two of the more salient findings were that daydreaming did not increase in the oldest ages, and that among BLSA men at any age daydreaming did not concentrate on the weird, outlandish, or improbable.

Daydreaming about the past. The common belief that old people spend much time daydreaming about the past is contradicted by a BLSA study of the past, present, and future settings of the daydreams of 1100 men and women aged 17 to 92 years (Giambra, 1977a). Using the three scales of the IPI that measure past, present, and future orientations, the study found no correlation between age and daydreaming about the past. Furthermore, temporal orientations were, with few exceptions, nearly the same in all age groups.

The relation between daydreaming and temperament. Reviews of studies in college students by other investigators have shown a connection between daydreaming characteristics and such measures of temperament as thoughtfulness, objectivity, masculinity, and emotional stability. Since Giambra (1974, 1977–78) had demonstrated that certain daydreaming characteristics in males are related to age, another BLSA investigation was conducted to determine how age may alter the relation (Giambra, 1977b).

The sample consisted of 170 males in six age groups: 24–34 (n = 17), 35–44 (n = 16), 45–54 (n = 28), 55–64 (n = 45), 65–74 (n = 36), and 75–91 years (n = 28). Aspects of daydreaming and related imaginal processes were determined by the IPI. Temperament traits were measured by scales from the Guilford-Zimmerman Temperament Survey (GZTS). A factor-analytic approach was used. Only one daydreaming-temperament factor—"neurotic-anxious absorption in daydreaming"—was age-related, and the relation was negative; older individuals had lower scores. Subsequent longitudinal analyses (see Chapter VI) demonstrate that this cross-sectional difference does not represent age change.

Sex differences in daydreaming. A study was subsequently undertaken to examine the daydreaming characteristics of a comparable sample of women of varying age (Giambra, 1979–80).

A total of 773 women (not drawn from the BLSA population) were divided into 12 age groups ranging from 17 to 92 years. The men were the sample from the BLSA used by Giambra (1977–78). The women were of the middle and upper-middle classes. Aspects of daydreaming and related mental activity were determined by the IPI. The frequencies of daydreaming about all subjects except problem solving decreased with age. Females reported higher frequencies than males of daydreaming and nightdreaming, as well as more problem-solving dreams. Females also reported lower frequencies of sexual, bizarre-improbable, heroic, and achievement-oriented daydreams. Most sex differences persisted over the life span, and the male-female difference in sexual daydreams increased with age. Except for sexual daydreams in males 17 to 29 years of age, daydreams in both sexes were primarily concerned with problem solving, which remained at a high level across the life span. After age 40, achievement-oriented daydreams were more frequent in females than in males. Females at all ages when contrasted with males reported more interpersonal curiosity. Males reported more curiosity than did females about things. An unexpected finding was that males were curious about things and people to the same degree. The results were interpreted as supporting the assumption that daydreaming is a way of dealing with current concerns or of solving problems.

Sexual daydreams and age. Retrospective reports of male sexual daydreaming and their relation to three behavioral aspects of sexual vigor were analyzed over the adult life span (Giambra and Martin, 1977). The three characteristics were frequency of coitus during the early years of marriage, number of partners before, during, and after marriage, and amount of sexual activity between 20 and 40 years of age. A total of 277 men from 24 to 91 years of age participated. The men were divided into six age groups: 24–34 (n = 23), 35–44 (n = 27), 45–54 (n = 56), 55–64 (n = 83), 65–74 (n = 48), and 75–91 years (n = 40).

Sexual daydreams declined in frequency and intensity with age and virtually disappeared after age 65. For men aged 24 to 64 years, scores on the sexual daydreaming scale of the IPI correlated with each of the three characteristics of sexual vigor. Men who had more frequent sexual daydreams tended to be more sexually active.

4. The Marital and Sexual History

In an introductory paper, in which the methods used in interviews to obtain marital and sexual information were described in detail, 603 BLSA men aged 20 to 79 were divided into age groups for a variety of comparisons with age (Martin, 1975). Across age groups numerous gradients were in evidence, some of which reflected the past experiences of different cohorts while others were accounted for by the effects of age.

The influence of birth cohort was apparent in the fact that, with increasing age at interview, proportionately more subjects reported early farm residence, late ages at onset of initial petting, coitus, and marriage, coital abstinence before marriage, and coital experience confined to a single partner. An aging effect was found in the gradients observed for a number of other variables: a history of marital dissolution and remarriage, the loss of erotic reactions to certain visual stimuli, an increase in the length of time during which abstention from sexual activity failed to produce discomfort, a reduction in kind and frequency of sexual expression, and recurrent erectile failure. The proportion of subjects who responded affirmatively when asked if they might want a restoration of sexual vigor, if this were possible, proved surprisingly small (33%) and showed no consistent trend with age. In view of the importance males are generally believed to place on sexual activity, it had been anticipated that with advancing age the prospect of renewed sexual vigor would have an increasing appeal.

Age, marital status, potency, and physical and emotional health are known to affect the frequency with which males engage in sexual activity. It is unclear, however, why males of similar age vary widely in frequency of sexual expression and why male sexual vigor declines with age. An analysis of the data obtained from 628 respondents was undertaken in an effort to identify correlates of sexual frequency while holding age constant (Martin, 1977). The hypothesis was that such correlates might identify factors or mechanisms that contribute to the loss of sexual vigor that sometimes accompanies aging.

Respondents were divided into five-year age groups, ranging from 25–29 years to 80–84 years at report. Mean and median weekly frequencies of coital and total sexual activity were computed from the frequencies reported for the five-year interval immediately preceding interview. Mean frequencies of total activity showed a slight increase from 20 to 34 years of age, but then declined steadily until age 65. The lack of a further decrement from age 65 to 79 was attributed to the above-average health of the older subjects. Coitus comprised an increasing proportion of total activity to age 34 and thereafter constituted 80% to 90% of total sexual activity.

A number of subjectively evaluated variables, most derived from the interview, proved unrelated to level of sexual activity irrespective of age. These included subject's religiosity, parental religiosity, economic status of the parental home, interpersonal relations within the parental home, number of times intoxicated, amount of physical activity, and health complaints on the Cornell Medical Index. Age at marriage was also independent of sexual frequency. However, total sexual frequencies at ages 20 to 39 were found to be significantly related to age at first coitus, number of coital partners, and maximum number of coital events reported for any single week of marriage, none of which appeared as correlates of total sexual frequency at older ages. On the other hand, the amount of sexual activity reported as having occurred between 20 and 39 years of age proved to be highly related to current rates of total sexual activity for all age groups past 40 years. Respondents had thus tended to maintain relatively high or low rates of total activity over most of their lives.

Physical and physiological variables that were found to be independent of current levels of activity at all ages included height, weight, lean body weight, percentage of body fat, size of prostate, hematocrit, hemoglobin, triglycerides, creatinine excretion, basal pulse rate, basal systolic and diastolic pressures, vital capacity, and forced expiratory volume at one second. Four variables appeared, however, as significant correlates of frequency of sexual activity at ages 65 to 79: chest circumference, maximum breathing capacity, and basal oxygen consumption as positive correlates, and serum cholesterol as a negative correlate. The small magnitude of the correlations, on the other hand, supported the conclusion that in persons of reasonably good health physical fitness is of minor importance to the maintenance of sexual vigor.

The problem addressed in a more recent paper (Martin, 1981) was to determine why some older men are more or less sexually active than others of comparable age. The study of 188 respondents 60 to 79 years of age provided a number of insights into the nature of male sexuality and served to generate several hypotheses concerning the factors that sustain sexual motivation in the later years of life.

Data analysis was limited to respondents who had been married throughout the year preceding the interview. Subjects were stratified into five-year age groups before being subdivided into categories of sexually least active (n = 63), moderately active (n = 63), and most active (n = 62), according to the quantity of sexual activity reported for the year prior to interview. Comparisons between least active and most active subjects revealed no important differences in occupational status, education, age at marriage, times married, number of years married before age 60, age of wife, or current marital adjustment. Nor were significant differences found between these groups in age at first coitus, number of coital partners, attitudes concerning coital and masturbatory activity, or perceived sexual attractiveness of wives.

Other variables, however, emerged as significant correlates. All four measures descriptive of past levels of sexual activity proved to be highly related to current levels of activity. This finding is consistent with the hypothesis that individual differences in level of sexual activity before middle age tend to be maintained past middle age, and that the persistence of these differences into old age accounts for a large portion of the variation observed in current frequencies.

An important conclusion from these findings is that men who were most sexually active in their 70s had also been highly active sexually in their 20s. The behavioral continuity is similar to the continuity of major personality dispositions.

The study also investigated sexual potency. Of the 88 men classified as less than fully potent at report, only 10% stated that they had sought medical advice for their condition, although none complained of fear of failure or of being unable to live up to some desirable standard of sexual performance. Moreover, questions about marriage and sexual experience, episodes of acute anxiety, and instances of resort to professional help for personal or sexual problems consistently failed to reveal evidence of the kinds of marital conflict, personal grievances, or expressions of dissatisfaction one would expect to find were emotional factors of critical importance for lack of potency. The vast majority of respondents appeared to be functioning at a level commensurate with their feelings of desire; because few were lacking in other resources for maintaining self-esteem, their condition failed to produce the emotional trauma that is often encountered in clinical practice.

5. Activities and Attitudes

A study was undertaken to measure the activities and attitudes of BLSA participants in order to compare the distributions of scores at different ages (Stone and Norris, 1966). The subjects were 463 men aged 20 to 99 years, 151 of whom were over 60. The Chicago Activity and Attitude Inventories of Burgess, Cavan, and Havighurst (Cavan et al., 1949) were administered to each participant. The scale is made up of three parts: background, including general information about the participant and his earlier life, an activity inventory, and an attitude inventory. The activity inventory, administered to all 463 men, contains five groups of statements dealing with leisure-time activities, religious activities, intimate contacts, security, and health. The attitude inventory, administered to 450 of the men, deals with the personal aspects of adjustment and contains eight groups of statements dealing with health, friends, family, work, happiness, economic security, religion, and feelings of usefulness.

No significant relations between activity or attitude scores and chronological age were found. In contrast with a previous study by other investigators (Mason, 1954) the older men in this group reported as much participation in and satisfaction with activities and relationships as the younger men. The contrasting findings were explained on the basis of differences in the study populations. The earlier study had compared a group of low-income institutionalized subjects, whose mean age was 74.2 years, with a group of high-income non-institutionalized subjects whose mean age was 70.2 years and a group of low-income non-institutionalized subjects whose mean age was 39.5; it found impressive differences in attitudes associated with age. In the BLSA study, the subjects were volunteers, were generally very well educated, and were working in or retired from high-level occupations. Most of the BLSA participants who have retired continue in activities that enhance their self esteem and allow them to feel that they are useful members of society. It is thus likely that the homogeneity of activity and attitude scores across age groups in the BLSA sample is a result of other similarities between the old and young participants.

6. Coping with Stressful Events

In a cross-sectional study of mechanisms used to cope with stressful life events (McCrae, 1982a), subjects from the Augmented BLSA Sample (participants plus their spouses) were administered questionnaires to measure coping behaviors and styles. A total of 255 men and women completed a Coping Questionnaire, and 150 completed a Coping Self-Interview. The Coping Questionnaire required subjects to indicate which of 118 ways of coping they had used in response to a recent life event selected by the investigator. The items were taken from the Lazarus "Ways of Coping" scale, together with fifty new items developed from a review of the literature. In addition, subjects indicated which ways of coping had been most helpful in their solutions of the problem and in making them feel better. The Coping Self-Interview required subjects to select three recent events—a challenge, a threat, and a loss—and to indicate which of 50 ways of coping they had used for each; whether the method had helped them solve the problem; and whether it had made them feel better.

Factor analysis of the 118 items in the Coping Questionnaire led to the identification of 28 coping mechanisms. Analysis of one step at a time found that restraint, rational action, expression of feelings, and positive thinking were the mechanisms most frequently used, whereas self-blame, intellectual denial, passivity, sedation, and hostile reaction were least frequently used. Although nearly half the mechanisms showed age differences in response to a recent life event, the kinds of events old people typically faced were systematically different from those young people faced; older men and women were more likely to have experienced a threatening event, younger individuals a challenge. The differences in type of stress were strongly related to the choice of coping strategy.

When statistical controls for type of event were used, eight mechanisms showed age differences. In order to replicate these findings, analyses of the Coping Self-Interview treated age and type of event as independent. Two of the eight age differences found on the Coping Questionnaire were replicated on the Coping Self-Interview: Older people (50–64 and 65–89 yr) were less likely than younger ones (21–49 yr) to use hostile reaction and escapist fantasy. These cross-sectional differences tend to support the work of Vaillant and to contradict the notion that older individuals are prone to the use of primitive and immature defenses. In addition, the many age differences related to the type of event testify to the fact that older individuals are able to adapt their coping behavior to the changing situational demands of their stage in the life span. Older men and women do not rigidly maintain habits of coping that, although appropriate in youth, have outlived their usefulness. Instead, as stresses change, so do coping responses. This finding was replicated in the second study, where non-significant age-by-type-of-event interactions indicated that older people showed the same differentiated and flexible responsiveness as younger ones to different types of stress.

CHAPTER VI

Longitudinal Studies of Aging

INTRODUCTION

There is not much evidence to support the concept of a general "aging process" that controls both physiological and psychological aspects of aging in an individual. It seems probable that aging, at least as we know it in humans, is rather the result of the interplay of many specific characteristics than a single process that regulates physiological and psychological functions. Aging is thus a highly individual phenomenon that can be characterized only by repeated observations in the same individual. This chapter summarizes the results of the Baltimore Longitudinal Study of Aging (BLSA) longitudinal analyses that have been completed and, in most instances, published. It must be looked upon as a progress report illustrating some of the findings about aging that have stemmed from longitudinal analysis.

Very few subjects have been followed over the full 25 years' existence of the BLSA, since new subjects have been added continually. Long periods of observation have, for example, produced five or more observation points on some variables for 667 subjects. Nevertheless, because of high variability of many of the measurements in different subjects, some longitudinal analyses have not yielded estimates of age trends in individual subjects with a desirable degree of precision.

The studies that are based on longitudinal analysis fall into five categories, for each of which a general summary has been prepared. The categories include a) studies that illustrate different conceptual approaches; b) analysis of anthropometric data on body size and composition; c) studies of physiological performance (basal metabolism, kidney function, blood cholesterol, and cardiovascular performance); d) studies of cognitive performance (vigilance, problem solving, learning, visual memory, the relation of temperament to visual retention, and the relation of hypertension to intellectual performance); and e) studies of personality characteristics (stability of personality, adjustment to aging, somatic complaints and neuroticism, hypertension and coronary disease in relation to personality, and personality and the life course).

Because of the many details that must be considered in longitudinal data analyses, we have chosen to reproduce in full all the publications from the Gerontology Research Center (GRC) that have used longitudinal analysis (see Appendix).

CONCEPTUAL CONSIDERATIONS IN THE IDENTIFICATION OF AGE CHANGES

By their very nature, longitudinal assessments of individual age changes commence with the acquisition of cross-sectional observations. Many of the studies reported in this chapter have been reported as cross-sectional investigations in Chapter V.

As Chapter I has noted, inferences of age changes from cross-sectional data fail to take into account birth-cohort effects. An example of the false impression that may arise from cross-sectional analysis of age differences is provided by our own experience with an assessment of the effect of age on body height and weight. Changes in these characteristics were examined in individual subjects over an interval of two to eight years during which three to eight measurements were made (*Shock, 1972*). The linear regression of height and weight on age was calculated for each subject on whom at least three measurements were available. Mean values for the slope of the individual regressions of height and weight on age were calculated for each subject within each age decade.

The average age decrement in body height, based on individually determined slopes, was similar to that derived by cross-sectional analysis. The results of cross-sectional analysis of body weight, however, differed from those based on longitudinal analysis of observations made on individual subjects. Cross-sectional analysis suggested a gradual decrement in body weight beginning with the 25-34 age group and continuing to age 80. Longitudinal analysis, on the other hand, demonstrated that, on the average, subjects 50 years of age or younger were gaining weight, while subjects 55 years and older were losing.

Early attempts to achieve an accurate longitudinal characterization of an individual's change in performance proved premature, since estimates of regressions on age based on only three observation points were highly variable and unreliable. Furthermore, practice or first-visit effects in some of the physiological variables greatly distorted the regressions on age when these first observations were included. For example, although three sessions were given to each subject in order to train him in performing the maximum breathing test (*Norris et al., 1964*), average values obtained on the second visit, 12 to 24 months after the first, were systematically higher than first-visit values, despite the fact that the average values fell with age and that values obtained on the third and subsequent visits decreased progressively. If the low initial value was used in the calculation of the age-regression coefficient, the resultant slopes did not accurately portray the changes in performance. None of the other pulmonary-function tests showed this practice effect.

The design of a longitudinal study requires the calculation or estimation of a number of statistical variables for each test. An extensive series of measurements of a variable serves as a data bank that can be used for determination of the sample size, the number of tests, the intervals between tests, and thus the total duration, required to characterize an individual's age changes in that variable with a specified reliability (*Schlesselman, 1973a,b*). These calculations require preliminary data to provide an estimate, based on cross-sectional data, of the average regression of the variable on age and of the total variance shown by the variable over the adult age span.

Schlesselman developed formulae and tables offering various combinations of study duration and number of measurements that yield the calculated value of "omega," a statistic calculated for each variable from estimated values of the slope of its regression on age, the standard error of estimate, and the selected degree of confidence in the computed individual slopes. For example, for an omega value of 0.20, three tests of kidney function distributed over a seven-year period provide as precise an estimate of the slope of individual age regressions as 30 observations made in three years. For a study of fixed duration, it is thus possible to establish the precision of the determination of the age regression that follows specified numbers of testings.

The longitudinal study of renal function (*Rowe et al., 1976b*) was based on the calculation of the regression of creatinine clearance on age in individual subjects. Although the *average* values for the individual slopes agreed surprisingly well with the average regression on age based on cross-sectional data, the study concluded that in order to characterize accurately the rate of age changes in creatinine clearance *in individuals,* serial observations over a period of at least 18 years may be required.

The study by *V.K. Elahi et al. (1983),* based on seven-day dietary diaries, illustrates the attempt to separate the effects of period and birth cohort from those of age. The nutrients considered were calories, protein, carbohydrate, fat, saturated fatty acids (SFA), polyunsaturated fatty acids (PFA), and cholesterol. Of the 12 indices derived from these variables, only three failed to show age effects: the percentage of calories derived from protein, intake of PFA, and P/S ratio. With the exception of carbohydrate, which rose, the intake of all of these nutrients declined with age. Birth-cohort effects were observed for none.

That period effects were also present over the duration of the study (1961–1975) was indicated by the systematic decrease in the absolute level of all the nutrients with the exception of PFA, which increased. The largest and most consistent period effects were an increase in the ratio of PFA to SFA and a decrease in cholesterol intake.

Douglas and Arenberg (1978) also attempted to separate age changes from cohort and period differences exhibited in responses to the Guilford-Zimmerman Temperament Survey (GZTS). Three hundred thirty-six subjects were measured on two occasions separated by an interval of seven years. To make possible cross-sequential and time-sequential comparisons, another 310 subjects were also tested for the first time when the original group were retested. The longitudinal analysis showed a significant age decline in Masculinity scores at all ages. The magnitude of the change increased with age. In contrast, General Activity scores declined only after age 50. Other scales (Thoughtfulness, Personal Relations, and Friendliness) showed declines from the first to the second administrations, but the magnitude of the change was not systematically related to age. The declines in General Activity and in Masculinity appeared to be age changes; the declines in Thoughtfulness and Personal Relations appeared to be related to changes specific to the period spanned by the two times of measurement as indicated by the differences within age groups in the time-sequential analysis; and the declines in Friendliness appeared to be attributable to period drift over a long time. In addition, later-born cohorts were lower in Restraint and higher in Ascendance than early-born cohorts.

Arenberg (1982a) devised a modification of the sequential strategies to explore further non-aging effects suggested by previous analyses of the Benton Visual Retention Test. Regression analysis, in which calendar time is a continuous variable, was used in an attempt to assess the magnitude of a period effect and to determine whether that effect could account for the within-cohort declines in performance. In four of the five oldest birth cohorts, declines were too large to be accounted for by the small period effects found for all age groups. Some portion of those declines was thus interpreted as representing the effect of aging.

BODY SIZE AND COMPOSITION

1. Fat Redistribution and Changing Body Dimensions

Cross-sectional population studies, as well as casual observation, indicate that the size, shape, and proportions of the human body change throughout adult life. Although all major tissues of the body change considerably with age, the changes in external size and body proportions occur because of changes in the relations among fat, lean tissue, and bone. The effect is most clearly evident in age-changes in subcutaneous fat distribution.

In a study that used both longitudinal and cross-sectional analysis, *Borkan and Norris (1977)* measured changes in subcutaneous fat distribution in relation to changes in other body compartments. The investigators compared various diameters and subcutaneous fat thickness to determine the role of fat, in relation to lean tissue and bone, in altering external dimensions.

The study population consisted of 699 BLSA participants aged 20 to 92 years. A total of 971 soft-tissue radiographs of these individuals, taken between 1958 and 1973, was available for analysis. Each radiograph contains views of seven body sites on the trunk and limbs. Because radiographic differentiation between skin and fat is difficult, measurements designated "fat" customarily include skin. A precision caliper was used to measure fat on radiographs of the trunk (at bony landmarks such as the top of the greater trochanter) and of the calf and forearm (at the widest parts). Reproducibility was tested by remeasuring a sample of 20 films; the average correlation between the paired measurements over the various sites was 0.95.

Other data used included body circumferences at sites corresponding to each radiographic fat measurement, height, weight, age, and biochemically assessed extracellular water and total body water. Values for extracellular and total body water were used to calculate the percentage of body fat. Total body water and extracellular water were assessed from the dilution curves for the distribution of standard intravenous doses of antipyrine and sodium thiocyanate. Extracellular water was subtracted from total body water to derive an estimate of intracellular water in the body.

Two separate data sets, one cross-sectional and one longitudinal, were used in the analysis. The cross-sectional sample, consisting of first-visit data for 699 individuals, was divided into ten-year age groups between the ages of 25 and 84. The longitudinal sample consisted of the first two of three radiographs of 234 participants (average interval between first and second films was 6.8 yr). Increments per year were calculated for all variables for each individual in the longitudinal sample. Each individual was placed in one of the ten-year categories of the cross-sectional sample based on his average age between the two test visits. Since third and fourth radiographs were available for very few individuals, they were not included in the longitudinal analysis.

Cross-sectional analysis showed that average values for both height and weight decline steadily after age 30. Total body water as a percentage of body weight declines in younger adults but becomes constant after age 55. Although the proportion of fat in relation to body weight increases with age, the calculated weight of fat changes little through the adult life span. The conclusion that the decline in body weight with age is due to loss of fat-free tissue is further supported by an age-dependent decline in percentage of intracellular water. The data on various body sites also gave evidence of a decline in fat-free tissue with age. Fat-thickness measurements of the trunk, particularly at the ilium and lower thorax, had the highest correlations with body weight. The marked reduction in forearm diameter after age 60 is due not to changes in fat thickness but to a loss of muscle tissue (Fig. V.5).

Longitudinal analysis showed weight increments in the three youngest age groups (25–34, 35–44, and 45–54 yr) and decrements between the ages of 55 and 74. The oldest individuals showed a marked increase in weight, possibly as a result of differential survival that results in unusually good health in very old individuals. Height decrements accelerated with age. Among measures of lean tissue, calculated fat-free weight declined in all age groups except the oldest. Body-fat increments, which were highly positive in the first age group, decreased with age, and the oldest group showed a marked decline. Decrements were observed in measures of arm and calf fat in all age groups. In the middle trunk, fat lateral to the lowest rib showed a decline through adulthood. In contrast to the cross-sectional data, which showed intervals of increase and decrease in fat lateral to the anterior-superior iliac spine, longitudinal data showed positive increments in the groups for which sufficient data were available (all groups except the youngest and the oldest). There were too few measurements at the greater trochanter to allow calculation of change.

The longitudinal data reveal that sites in close anatomical proximity behave differently during adulthood. While fat in the extremities remains relatively stable in the cross-sectional data, the longitudinal data show that it declines through adulthood, the middle trunk undergoing a net loss and the lower trunk a gain in fat. This shows that subcutaneous fat is in a dynamic state throughout adult life. The change is most striking in the longitudinal data, which show that subcutaneous fat in some sites may actually decline during the years in which overall body weight and fat increase.

These findings suggest that, at least during old age, when all subcutaneous fat in most sites is either declining or unchanged in thickness, total fat remains the same. The implication is that internal fat is accumulating while subcutaneous fat is decreasing with aging.

2. Height and Weight

Studies of the loss of height and weight with age have been described above, under "Conceptual Considerations in the Identification of Age Changes." It was found that height showed a decrement with age, both in age groups and in individuals. On the other hand, calculation of age regression of body weight in individual subjects showed that for the decades 25–34, 35–44, and 45–54 the regressions on age were positive even though the average regression calculated from cross-sectional data was negative.

3. Joint Degeneration in Osteoarthritis of the Hand

Longitudinal anthropometric measurements based on radiographs were used by *Plato and Norris (1979b)* in their study of the rate of change in joint degeneration in osteoarthritis of the hand. The authors studied 478 BLSA participants between 21 and 97 years old who were grouped in four categories according to their ages at the time of their first x-ray. Group A included 107 individuals below the age of 40 ($\bar{x}$ age = 33.2 yr); group B 175 participants aged 40-54 ($\bar{x}$ age = 46.9 yr); group C 129 participants aged 55–69 ($\bar{x}$ age = 61.8 yr); and group D 67 individuals who were x-rayed for the first time when they were 70+ years old ($\bar{x}$ age = 74.1 yr).

The x-rays were graded according to published standards that recognize five grades of severity of osteoarthritis (grades 0 and 1 are considered normal, grade 4 severe). Osteoarthritis in the distal and proximal interphalangeal joints was evaluated separately by the DH, PH, and IH scoring system. The DH score represents the highest osteoarthritic grade among the five distal interphalangeal joints, the PH score the highest grade among the four proximal interphalangeal joints plus the metacarpophalangeal joint of the thumb. The IH is the highest osteoarthritic grade observed in any of the interphalangeal or metacarpophalangeal joints of the hand. All radiographs were evaluated by the same investigator to eliminate variability.

The results showed that joint degeneration from osteoarthritis is a relatively slow process. The maximum rate of degeneration was seen in the distal interphalangeal joints, where the average increase was about one grade per individual in the interval of 12 to 16 years between x-rays. The rate of degeneration in the proximal interphalangeal joints was much slower.

The progress of degeneration in distal interphalangeal joints of individuals (longitudinally evaluated) was very similar to that seen in an earlier cross-sectional study by the same authors (Plato and Norris, 1979a).

The rate of change in the osteoarthritic grade of individual hands agreed closely with that of their distal interphalangeal joints. This further supports the conclusion of the authors' earlier cross-sectional study (Plato and Norris, 1979a) that what has been referred to as an osteoarthritic grade of the hand of an individual may actually be the highest grade among the distal interphalangeal joints.

PHYSIOLOGICAL PERFORMANCE

1. Basal Metabolism

Basal metabolic rate (BMR) has traditionally been expressed in relation to body-surface area, which is calculated from height and weight by a standard formula. The use of body size as a reference, however, assumes that all tissues contribute equally to body metabolism. In reality, adipose, bone, and connective tissues have very low oxygen demands and contribute very little in comparison with other tissues to the bulk of the whole-body basal oxygen consumption. Early in this century, "active tissue mass"—the estimated weight of tissues actively consuming oxygen—was proposed as a more suitable reference for the BMR, but the difficulty of measuring it accurately has made the approach impractical.

A BLSA cross-sectional analysis of aging by Tzankoff and Norris (1977) confirmed results reported by Shock et al. (1963) showing that age-related decrements of whole-body basal oxygen consumption in adult men were largely attributable to concurrent decrements in the mass of metabolically active (creatinine-producing) skeletal muscle. The approach was advantageous in that it was not necessary to account for either the deposition or the loss of fat, and the questionable reference to body size was avoided. The authors later extended these findings in a longitudinal study of 355 BLSA participants in whom five or more paired determinations of whole-body oxygen consumption and muscle mass were obtained over a mean period of 10.7 years *(Tzankoff and Norris, 1978)*.

Nonmuscle oxygen demand and muscle oxygen demand in the basal state were calculated from paired values of whole-body oxygen consumption and 24-hour creatinine excretion by the method of Tzankoff and Norris (1977). Each subject's data were represented by a slope obtained by least-squares linear regression of each variable on age. They were then grouped in age decades. The overall mean rate of change, –0.82 ml O_2/min per year, was similar to that calculated from cross-sectional data. Mean slopes of whole-body oxygen consumption, summarized by age-decade groups, were all negative, not very different from one another, and consistent with the cross-sectional trend. Nonmuscle oxygen consumption increased for the six older groups, in contrast with the lack of a cross-sectional age effect in skeletal muscle.

Among these subjects, 48 had died shortly after their last measurements ($\bar{x}$ interval = 1.9 yr), 46 of them from cancer or cardiovascular disease. As a group, the decedents had had significantly higher mean slopes for nonmuscle oxygen requirement than the average. In the absence of these diseases, muscle-mass decrement and the concomitant decrease in oxygen requirement accounted for age decrements in whole-body oxygen requirement. In the subjects who died, however, the overall decline had been slowed or even reversed by gradual increases in nonmuscle oxygen requirement during nearly the entire decade before death. These findings led the authors to qualify a conclusion of the earlier cross-sectional study that loss of muscle mass is responsible for all the decline in whole-body oxygen requirement with age. Men nearing terminal age were similar to the survivors in that they lost skeletal muscle at comparable rates, but they differed from the survivors in that their decline in muscle oxygen requirement was offset by gradual increases in nonmuscle oxygen requirement. The result was a lower rate of decline in their whole-body oxygen consumption. In longitudinal studies that include only two points, one early and one very late, this effect could be mistaken for a slower rate of aging.

2. Renal Function

All participants in the BLSA studied between July 1, 1961, and June 30, 1971, were included in a longitudinal and cross-sectional study of creatinine clearance *(Rowe et al., 1976b)*. More than 3300 creatinine clearances were obtained in 884 volunteers ranging in age from 17 through 96 years. On the basis of clinical examinations, subjects were placed in categories indicating the presence of specific diseases or medications that might alter the glomerular filtration rate. Only subjects who were not in these categories, a total of 548 individuals, were included in the data analysis.

Cross-sectional analysis showed a highly significant reduction in creatinine clearance with advancing age in healthy subjects. The data when grouped by decades suggested that creatinine clearance remained stable until age 34 and thereafter declined, the rate of decline increasing with each decade after age 65. In the cross-sectional analysis, however, this tendency toward an increasing rate of decline in the very old subjects proved not to be statistically significant.

The longitudinal analysis consisted of the calculation of the regression of creatinine clearance on age for each subject on whom five or more tests had been completed. The means of the individual slopes in the longitudinal analysis were similar to the mean differences in the cross-sectional analysis. Regression analysis of individual slopes by age indicated a minimally significant acceleration in the rate of decline in renal function with age after age 65. In view of the detailed screening of the subjects to eliminate those who might possibly have suffered from overt or occult renal disease, the authors concluded that the decrements represent true renal aging rather than the development of renal disease.

Shock et al. (1979) analyzed the longitudinal data from 398 BLSA subjects, aged 25 to 100 years, who were tested for 24-hour creatinine clearance five or more times over a period of ten years. They found that the correlation of individual regression slopes with age was highly significant, i.e., the rate of fall in creatinine clearance increased with age over the range of 25 to 90 years. Average slopes for 20-year age groups became increasingly negative, from -0.26 ml/min·1.73 m^2 per year for the 20–39 age group to -1.51 for subjects aged 80 to 100 years. This relation could not be demonstrated from cross-sectional data.

Although the standard error of individual regression coefficients based on as few as five measurements may be large, eight subjects were identified with statistically significant positive coefficients (i.e., their creatinine clearance improved with age). Longitudinal observations can obviously identify individual subjects who deviate markedly from the average pattern of age changes derived from cross-sectional data.

The same study yielded five or more observations on 59 subjects who died after the age of 55. Mean values for creatinine clearance and individual regression slopes for this group were compared with values based on a group of 106 subjects over the age of 55 who were still living after ten years. The mean ages of the two groups did not differ significantly. Although examination of death certificates showed that renal disease was not recorded as a cause of death for any of the 59 subjects, the individuals in this group had exhibited lower creatinine clearances and faster rates of decline in clearance than subjects of comparable age who were still alive.

The longitudinal observations illustrated clearly the individual differences that exist in the rates of change of these variables. Some individuals follow the pattern predicted from cross-sectional analyses (i.e., a gradual fall in clearance with age). However, some of the normal subjects showed a pattern of increased creatinine clearance with advancing age—at least to age 70. This is an important finding that bears further longitudinal exploration.

3. Serum Cholesterol

A decline in cardiovascular mortality in the past decade has been noted by several authors. The decline has coincided with public-health efforts to control such risk factors for coronary artery disease (CAD) as blood pressure, cigarette smoking, physical-activity levels, obesity, and serum cholesterol. Independent cross-sectional studies have provided evidence that serum-cholesterol levels have been dropping in recent years, and similar changes have been suggested by preliminary longitudinal studies. The trend to lower cholesterol levels was confirmed in a detailed study by *Hershcopf et al. (1982)* of longitudinal changes in serum-cholesterol levels in BLSA participants over a period of 14 years.

Serum cholesterol, height, and weight are among the variables that have been followed in the BLSA since its inception. In addition, dietary and exercise histories have been maintained in subgroups of this well-characterized population. It was thus possible to measure not only changes in serum cholesterol but also the effects on cholesterol levels of body weight, diet, and physical activity.

A total of 1011 males ranging in age from 17 to 102 had 5127 cholesterol determinations between 1958 and 1977. After data exclusions, 3088 cholesterol determinations in 783 participants were available to determine changes over the 14-year period from 1963 to 1977. As a measure of obesity, weight was corrected for height by use of the body-mass index defined as weight in kilograms divided by the square of the height in meters. Dietary intake was assessed from seven-day diaries kept by each participant under the guidance of nutritionists. Estimates of physical activity were obtained by interviews or questionnaires covering specific activities at home, work, or recreation, as well as variations in activities such as trips and seasonal sports.

Total daily energy expenditures were calculated for each subject by use of predetermined values for each activity (McGandy et al., 1966).

In the analyses of serum-cholesterol changes, the annual rate of change was computed as the slope of the regression line for each subject with three or more data points. In the analysis of secular change, each participant was represented by the mean of available determinations within each period. Simple linear correlations and regressions were used to assess the relation between individual performances.

Cross-sectional analysis revealed age differences in cholesterol data obtained between 1971 and 1977. Higher cholesterol values were found with increasing age from 25 to 64 years, but lower values were found in the 65–84-year-old group. When the data were examined as successive cross-sectional studies, serum-cholesterol values were fairly constant in the periods between 1964 and 1970 but dropped 6% in the periods between 1970 and 1972. Since then there has been little change. Because of the drop in serum cholesterol, the total span of the study was separated into two overlapping periods, Era 1 (1963 to 1971) and Era 2 (1969 to 1977). Within-age-group differences in serum cholesterol between Era 1 and Era 2 were found in all age decades studied.

Longitudinal analysis of serum-cholesterol values obtained within Era 1 showed that changes in cholesterol levels in individuals grouped by age followed the pattern seen in the cross-sectional analysis, i.e., cholesterol increased in the younger adult years and decreased in the later years. The drop in cholesterol levels between Era 1 and Era 2 seen in the comparisons within age groups was strikingly evident in the longitudinal analysis as well.

A shift in methodology was ruled out as an explanation of the period change in cholesterol values between Era 1 and Era 2, since re-analysis of stored frozen and lyophilized samples yielded values parallel to the original ones. When the effects of obesity, selected dietary constituents, and physical activity were examined as potential explanations, serum-cholesterol levels were not significantly correlated with levels of weight or body-mass index. Changes in weight, however, were significantly and positively correlated with changes in serum cholesterol. The fact that the study population did not as a whole experience a significant decline rules out weight changes as a source of the period drop in cholesterol levels. There were no significant correlations between the absolute value of any dietary variables examined and the absolute level of serum cholesterol. Although there were small but significant changes in most dietary constituents, only changes in caloric intake had significant positive correlation with changes in serum cholesterol. The small overall change in caloric intake, however, could explain less than 1 mg/dl of the average 11 mg/dl drop. There was no overall change in physical activity, and no significant correlations were found between either the level of or the change in physical activity and the level of or the change in serum cholesterol.

Since changes in fatness, caloric intake, and physical activity cannot fully explain the change in cholesterol levels between Era 1 and Era 2, other factors that must be involved should be examined in future studies.

4. Cardiovascular Function

Right-bundle branch block in apparently healthy men. Although the prevalence and incidence of several electrocardiographic (ECG) abnormalities clearly increase with age, the long-term prognostic implications are controversial primarily because the effects of aging and disease on the ECG have not been adequately separated. One such ECG abnormality is right-bundle branch block (RBBB). Most of the studies examining the long-term prognosis of this conduction defect are derived from hospital-based or military populations, neither of which is representative of the general community. Men from the BLSA provided a more reasonable population for such a study, and an age-matched control population made possible the separation of aging and disease processes.

Resting 12-lead ECGs identified 24 men with RBBB who had shown no evidence of cardiac disease on initial presentation. To determine their long-term cardiovascular prognosis, they were compared *(Fleg et al., 1983a)* with an age-matched control group; mean age on presentation with RBBB was 63 years. No differences in the prevalence of major coronary risk factors or obstructive lung disease were noted initially. Over the 8.4-year average follow-up period, the incidence of angina pectoris, myocardial infarction, cardiomegaly, congestive heart failure, high-grade heart block, or cardiac death did not differ between RBBB and control groups. Likewise, no differences in cardiothoracic ratio, mean blood pressure, maximal aerobic performance, or maximal exercise heart rate were found between groups on most recent examination. Left-axis deviation, however, occurred more commonly in RBBB subjects (46% vs. 15%, $p < .01$), and their mean resting heart rate was lower than that of controls. In addition, PR-interval prolongation of 40 msec or more over time was more frequent in men with RBBB than in controls (21% vs. 6%, $p < .03$). These findings suggest that RBBB in clinically healthy men is found primarily in older individuals and has no effect on long-term cardiovascular morbidity or mortality. The frequent association of RBBB with left-axis deviation, as well as the slower heart rate and the fact that the tendency for prolongation of atrioventricular conduction with time is greater in men with RBBB than in controls, suggests a primary disorder of cardiac conduction in the former.

Longitudinal cardiopulmonary chest x-ray changes in normal men. The standard chest x-ray is the most common radiographic procedure employed in medicine. Despite the ubiquity of the examination, most of the information on radiographic changes seen with advancing age has been derived from cross-sectional studies in populations not carefully screened for the presence of cardiopulmonary disease.

To determine the changes attributable to the aging process, we evaluated cardiovascular and pulmonary structures on two standard postero-anterior chest x-rays taken at least ten years apart ($\bar{x}$ interval = 16.9 yr) in 67 carefully screened healthy men *(Ensor et al., 1983)*. Only normotensive individuals with negative maximal-exercise ECGs and normal pulmonary-function tests were included in the survey.

The mean aortic knob diameter increased from 3.4 ± 0.6 cm to 3.8 ± 0.5 cm, enlarging in 78% of subjects. Although mean cardiothoracic ratio (CTR) increased from .405 ± .04 to .427 ± .04 overall, only 3% of men developed a CTR greater than .50, and none exceeded .51. Pulmonary abnormalities on initial chest x-ray consisted mainly of hyper-inflation (27%) and increased markings (19%), both of which doubled in prevalence during follow-up. Kerley B lines and enlarged pulmonary arteries, which were rare initially, increased four- to fivefold over time. Commonly accepted x-ray criteria suggested chronic obstructive lung disease in 21% of the final chest films, despite the absence of clinical or spirometric abnormalities. These data provide a framework by which chest radiographic changes with age can be interpreted. The most significant finding is that the frequently quoted value of .50 for the normal upper limit of CTR remains valid even in advanced age.

Cardiothoracic ratio. CTR—the ratio of the cardiac diameter (CD) to the thoracic diameter (TD) as determined from x-ray measurements—has been used to assess heart size. CTR values exceeding 50% have been regarded as evidence of abnormal enlargement of the heart. Cross-sectional studies on other populations have reported values exceeding 50% in 10% of older men and in 20% of older women. Although these studies excluded individuals with overt heart disease, rigorous screening for the presence of heart disease was not conducted.

The availability of standard six-foot postero-anterior chest x-rays on BLSA subjects made it possible to examine the effect of age on this ratio *(Potter et al., 1982)*. The study used 1124 chest x-rays from 243 men, aged 20 to 95 years, who had been followed for an average of 12.3 years (range of follow-up was 8 to 21 yr). Criteria for inclusion in the study were a) three or more technically adequate x-rays; b) eight or more years between first and last x-rays; c) age at time of last x-ray under 50 or over 60 years. Measurements of CD and TD, which were made blind to subject age and date of x-ray, were reproducible to 0.1 cm.

In the longitudinal analysis the rate of change in CD and TD was computed as the slope of the age-regression line for each subject estimated from three or more data points.

Cross-sectional analysis indicated a slight but significant rise in average values of CD up to age 75, whereas TD showed no change between the ages of 32–50 and 60–69 years but fell after age 70. The ratio CD/TD increased progressively from the 32–50-year group through the 80–95-year group.

Longitudinal analysis indicated that changes in CD, TD, and CTR in individuals generally followed the pattern suggested by the cross-sectional analysis: increasing CD and CTR after 50 years of age, and decreasing TD in the oldest groups.

Patients with diagnosed heart disease had higher values for CD and CD/TD ratio at all ages, but TD did not differ significantly from values for the normal controls.

For each of the 49 deceased subjects a surviving subject was selected who had been the same age at the time of first testing and whose cardiovascular status had been the same. Comparisons of mean slopes of CD and CTR in deceased individuals with those in survivors showed that individuals who had died with diagnosed heart disease had had significantly higher slopes than survivors. However, individuals who died without evidence of heart disease had slopes that did not differ from those of matched survivors.

COGNITIVE PERFORMANCE

1. Vigilance

"Vigilance" refers to a central process or state reflecting the individual's readiness to respond to specific infrequent and unpredictable stimuli. The primary index of vigilance is the proportion of the infrequent signals that is detected. The test consists of detecting and reporting instances when the pointer attached to a large blank circular clock face moves two intervals instead of the usual single interval. The full circle includes 100 intervals. In the course of the one-hour test, 23 double jumps occur at irregular intervals. A cross-sectional study (Surwillo and Quilter, 1964) showed that fewer signals were detected by an old than by a young group. The study also showed that middle-aged individuals had the fastest reaction time to infrequent signals, outperforming both old and young. The reduced proportion of stimuli detected and the slower reaction times demonstrated by elderly subjects suggest a reduction in the reactivity of the central nervous system late in life.

Thirty-three subjects from the original study were retested, by the same procedures, after an interval of 18 years (*Quilter et al., 1983*). In addition to the cross-sectional comparisons of the second-time data and the longitudinal analysis of change, measures of men who were 51 to 69 years of age at the time of first testing were compared with measures of men who had reached ages of 51 to 69 by the time of their second testing.

The cross-sectional analysis of the number of double jumps that were detected showed a significant age effect. Subjects 70 to 88 years old detected 58% of the double stimuli, while 51–69-year-olds detected 71%.

A significant longitudinal effect occurred in those who were 70 to 88 years old in 1980–1981; they detected fewer targets (58%) than they had in 1962–1964, while those 51 to 69 years old in 1980–1981 detected one percent more targets than in 1962–1964. Longitudinal analysis thus confirmed the cross-sectional findings that young and middle-aged groups perform equally well on vigilance, while individuals approaching the age of 70 show a decline.

The longitudinal results also showed that in 1980–1981 the 51–69-year-olds had faster reaction times than 18 years previously, while the 70–88-year-olds were slower. The faster reaction times in the middle-aged group confirm the original cross-sectional study. It should be noted that the men who were 51 to 69 in 1980–1981 were similar to the men 51 to 69 in 1962–1964 in both proportion of signals detected and reaction time. Instead of increasing progressively over the entire age span, reaction time to infrequent signals increases primarily after age 70.

The results of the combined cross-sectional and longitudinal analysis and comparison support the conclusion that differences in vigilance, although they are not manifest in performance until after the age of 70, are true aging effects and are not due to cohort differences.

2. Problem Solving

Few studies have measured changes in problem-solving performance with age. Studies published before 1974 were cross-sectional. They showed that older adults had substantial difficulty in solving logical problems, but that some tests, which measure the ability to solve concept-type problems, did not consistently reveal differences between young and old adults. *Arenberg (1974)* was the first to publish the results of a longitudinal study of the effects of aging on problem-solving performance. The aims of the study were to determine whether performance on logical problems changes with age and whether old subjects decline more than younger subjects. A sample of 300 BLSA participants ranging from 24 to 87 years of age was studied.

All problems were administered during a single half-day session for each subject. After instructions, which included a sample problem and a practice problem, the subjects were given the main set consisting of up to three problems of increasing complexity. The second session was scheduled a minimum of six years after the first ($\bar{x}$ interval = 6.7 yr). Of the 300 initial participants, 224 were available to take part in the second session.

Each input prior to achievement of a solution sequence could be evaluated as potentially informative or uninformative, depending on the pool of information available to the subject at that point. The number of uninformative inputs was the primary dependent measure. These inputs were further classified as overtly redundant, directly inferable, or indirectly inferable, in order to elucidate the age deficits in reasoning that were hypothesized.

Performance on the tests was analyzed both cross-sectionally and longitudinally. Cross-sectional results for all problems showed that, with increasing age, the proportion of subjects who achieved solutions decreased and uninformative inputs increased. Although the pattern of age differences varied somewhat, all three problems showed a decrease in proportion of successful solutions and a decline in effectiveness of solutions with age.

Longitudinal measures of change in both Problem 1 ($n = 193$) and Problem 2 ($n = 166$) showed a mean decline in performance only for the group that was over 70 at the first testing. For that group, the number of uninformative inputs increased primarily because of the greater number of overtly redundant inputs, even though memory demands were minimized and the entire written record of input-outcome events was always available to the subject. Thus the oldest subjects repeated inputs frequently, and the increase in repeated inputs accounted for a major portion of the decline in the effectiveness of their reasoning. These subjects also made attempts at solution sequences that were variations of earlier attempts but were ineffective because they did not make use of the information available from previous input-outcome events.

A follow-up analysis of mortality and reasoning performance for the 49 men in the initial group over 70 showed a relation between survival and success in solving Problem 1. Of the 36 men ($\bar{x}$ age = 74.3 yr) who solved the problem, six had died; of the 13 ($\bar{x}$ age = 76.0 yr) who failed to solve the problem, six had died. There had been no difference in vocabulary scores between the survivors and nonsurvivors.

The true age changes in this study were probably reduced as a result of several positive biases: Individuals who returned for the second test were a select subsample of the original group; most subjects who reached Problems 2 and 3 were superior performers; and only those who successfully solved a particular problem in both sessions could be included in the analysis of change in reasoning effectiveness. Although a substantial mean decline was found for the group over 70, even beyond age 75 the performance of some subjects had not declined from that observed six years earlier.

In 1967, a study of concept problem solving was initiated in the BLSA. Cross-sectional and longitudinal data collected from 1967 through 1979 were analyzed *(Arenberg, 1982b)*. The problems were designed to measure not only whether correct solutions were reached but how effectively each was solved.

There were twelve problems: four one-attribute (simple), four two-attribute with high initial information, and four two-attribute with low initial information. Each problem comprised four dimensions, and each dimension had two attributes. To make the problems easier to understand, they were described as attempts to identify the "poisoned" food or foods in a problem. In the "language" of poisoned foods, the attributes are foods, the concepts (solutions) are poisoned foods, subjects select meals, and each meal is designated as having caused an imaginary diner to "Die" (positive) or "Live" (negative), depending on whether it included the poisoned food or foods.

To minimize the memory component of the task, subjects were required to write each selection and its designation (provided by the experimenter). Each concept rule (simple, conjunctive, disjunctive) was explained thoroughly and was prominently displayed during the problem. The entire procedure was subject-paced.

During the first 13 years of this study, complete first-time data were collected for 751 men ranging in age from 20 to 87 years. Cross-sectional analysis of the number of correct solutions found a monotonic decrease in means from 10.4 for the men in their 20s to 5.4 for the men in their 80s.

By the end of 1979, 327 men had returned for a repetition of all 12 problems at least six years after the first session ($\bar{x}$ interval = 6.9 yr). Mean improvements were found for the younger age groups, but the men who had been in their 60s or 70s when first measured showed small mean declines in the number of correct solutions.

Estimates of age changes were calculated within birth cohorts by regression analysis of number correct (first session) and calendar time (date of session). These within-cohort estimates showed no change for the six ten-year cohorts born between 1897 and 1956; but for the earliest-born cohort (1887–1896) the estimate was –0.25 problems per year. The estimates of age change thus confirm the direct (longitudinal) measures of change, i.e., declines in performance only late in life.

Results using the effectiveness measures were similar to the results for number of correct solutions.

3. Learning

BLSA longitudinal data on age changes in learning ability, although largely preliminary, constitute perhaps the most extensive body of information on the subject. The data on performance in paired-associate and serial learning at two different anticipation intervals for each task were described by Arenberg and Robertson-Tchabo (1977) and *Arenberg (1983)*. Arenberg (1967b) had previously shown by cross-sectional analysis that small age differences existed before age 60 and larger age differences thereafter, particularly when only a short anticipation interval was used.

The BLSA data on age changes in learning are of two kinds: conventional longitudinal data based on repeated measures at least six years apart, and independent subsamples of the same birth cohorts measured during two separate periods. Six birth cohorts, based on dates of birth between 1885 and 1932, were studied in the early sample, which was measured between 1960 and 1964. The second subsample of each cohort comprised men who entered the study after the first sample, but data for only five of the six cohorts were available for the later sample, which was measured between 1968 and 1974. Comparison of independent subsamples born in the same periods (cross-sequential analysis) is advantageous in that it eliminates retest effects and the effects of noncomparability of learning material, but is disadvantageous in that age becomes a between-subjects variable and any precision attributable to the within-subjects design is lost. The advantage of the design lies in the argument that if repeated-measures data are corroborated by intra-cohort differences, the evidence for age changes is compelling.

The mean period between tests was 6.8 years for the 102 men assigned to the short anticipation interval (fast-pace condition) who had two valid measures on the paired-associate task. The youngest group (initially 30–38 yr) showed a small mean decrease in errors. The other two groups of men initially less than 54 years of age showed moderate mean increases. The three oldest groups showed larger mean increases in errors, the largest in the oldest group (initially 69–76 yr). Cross-sequential analysis found the smallest increases within birth cohorts in the three latest-born cohorts (1909 to 1932) and the largest increases in the two earliest-born cohorts (1893 to 1908), i.e., the oldest groups. Mean raw scores on the Vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS) were similar in the early (1960–1964) and the later (1968–1974) samples.

The mean interval between administrations was 6.7 years for the 111 men assigned to the slower pace who had two valid measures on paired-associate learning. The three youngest groups (initially 30–55 yr) showed the smallest mean declines, the two oldest (initially 61–74 yr) the largest. The smallest difference within birth cohorts was found in the youngest cohort (1925–1932). Large differences favoring the younger subsamples within birth cohorts were found for the two earliest-born cohorts (1893–1908), the oldest groups. Again, the mean WAIS Vocabulary raw scores for the early and later samples were quite similar.

For the 104 men who had two valid measures of serial learning at the faster pace, errors decreased, although the two oldest groups (initially 63–76 yr) showed mean increases. Comparison of early and late subsamples within birth cohorts showed the largest increases in errors in the two earliest-born (oldest) cohorts.

For the 102 men who had two valid serial-learning measures at the slow pace, errors increased. Again the largest decline in performance was found in the oldest group (initially 68–76 yr). All birth cohorts showed age differences; mean errors were consistently greater for the older subsample within each cohort. Although no clear pattern emerged between the magnitude of the mean age differences and the birth periods of the cohorts, the earliest-born cohort showed the largest age difference.

Thus, for both paired-associate and serial learning, age decrements in conventional longitudinal comparisons were found in the oldest age groups. In addition, comparisons within birth cohorts showed that in early-born cohorts older subsamples performed less effectively than younger subsamples. The evidence indicates a deficit in verbal learning in the later years of life.

4. Visual Memory

In a longitudinal study of age changes in nonverbal memory published by *Arenberg (1978)*, performance consisted of reproducing geometric designs from memory. In addition to cross-sectional data, the study included age changes based on conventional longitudinal data as well as on comparisons of independent samples from the same birth cohort measured at different times.

The subjects included every participant in the BLSA for whom a measurable initial performance during the period from 1960 to 1973 was available, a total of 857 men whose age range was 18 to 102 years. The total sample was divided into early, middle, and late subsamples based on the date of the initial measurement. The primary cross-sectional analysis was based on the early sample, the middle and late samples serving as replications. The primary longitudinal analysis was based on the re-examination of 268 men from the early sample, and a small longitudinal replication was based on the re-examination of 82 men from the middle sample. Age changes were also estimated in subsamples from the early and late samples on the basis of comparisons within birth cohorts (age differences between subgroups born at the same time but measured at different times).

All participants were given Form C of the Benton Revised Visual Retention Test, Administration A, usually on their first visit. Form E was administered on a subsequent visit at least six years after the first test. Errors were scored independently by two psychologists according to the test manual. At the same session the WAIS Vocabulary subtest was also administered.

For cross-sectional analysis, subjects from the early, middle, and late samples were classified in seven age groups. In all three samples the mean number of errors increased with age.

Although the magnitude of change increased with age in both longitudinal samples, statistical significance was not attained in the middle sample. Assessment of cumulative distributions of age changes for five age groups in the early sample demonstrated a striking difference between the oldest group and the other four: Fewer than half the men in each of the four youngest groups declined in performance, but more than two thirds of the group initially over 70 declined. The large mean increase in errors for the men in their 70s was attributable not to a few men who changed drastically but to a shift of the entire distribution in the direction of larger decrements in performance. Similarly, estimates of age changes based on comparisons within birth cohorts indicated age differences in all cohorts, the largest occurring in the earliest born.

In summary, increases in errors were small for the young groups, moderate for men in their 50s and 60s, and substantial for men over 70. WAIS Vocabulary measures for these same samples showed small cross-sectional differences favoring the older men, no overall longitudinal change but small deficits for the older participants, and small declines in estimates of age changes based on comparisons of independent samples born during the same period. In general, the results indicate age decrements in memory-for-design in men late in life, but small or no decrements in vocabulary for the same samples. No correlation was found in either longitudinal sample between change in memory performance and change in vocabulary score.

The picture is quite clear for memory for designs. Cross-sectional, longitudinal, cross-sequential, and time-sequential analyses provide consistent and strong evidence of age decrements late in life.

5. Temperament as a Predictor of Change in Visual Retention

Individual differences raise the important question whether predictors of an individual's decline in memory performance can be identified. In this case it might be possible to devise strategies for intervention in cognitive problems in the elderly. A preliminary BLSA study was therefore undertaken to examine temperamental traits as possible predictors of change in visual-memory performance *(Robertson-Tchabo et al., 1979)*.

The 52 subjects in the analysis included every BLSA participant 70 to 79 years of age at the time of the initial administration of the Benton Test for whom there was a second performance measure, as well as both initial and second administrations of the GZTS, which assesses General Activity, Restraint, Ascendance, Sociability, Emotional Stability, Objectivity, Friendliness, Thoughtfulness, Personal Relations, and Masculinity. Multiple regression analyses were carried out with the ten scales of the GZTS as independent variables and residualized change in Benton total errors as the dependent measure. (Change in Benton total errors represents the part of the second measure that is uncorrelated with or independent of the initial level of total errors.)

The results of the multiple regression analyses, together with the simple correlations, showed that the variables Objectivity, General Activity, Restraint, Masculinity, and Ascendance were important correlates and predictors of Benton performance.

Although the consistency with which some of the scales emerged as important predictors in this exploratory study is encouraging, the small sample size and the need to control for the effects of other variables, such as health factors, clearly require replication of the findings. Furthermore, the observed relation between personality traits and visual-memory performance in this study permits no conclusions about cause and effect or the underlying mechanisms.

6. Blood Pressure and Intelligence

Although the prevalence of elevated blood pressure in older people has led to suggestions that it may be partly responsible for decline in cognitive function with age, the literature on the subject is contradictory (Wilkie and Eisdorfer, 1971). Uncertainty arises in part from the complexity of the variables under investigation. Hypertension is not a single clinical entity; there are many kinds, and this diversity may influence any relation with intellectual performance. Most research has been conducted in clinic patients, who are often under medication that could influence intellectual performance. Different kinds of cognitive tests show different age effects: Visual memory, as the preceding section has shown, decreases significantly after age 70, while vocabulary does not in general decline and may even increase. Other frequent sources of inconsistency are small sample size (which makes it difficult to measure hypertensive effects that are small in comparison with the range of individual differences in intelligence) and faulty sampling procedures.

Longitudinal studies are needed to assess the long-term effects of increased blood pressure on intellectual performance, since serial observations in the same individuals over time permit the determination of the relations between antecedents and consequences. A BLSA longitudinal study by *Costa and Shock (1980)* used the Army Alpha test, a general intelligence test that includes some components of intelligence that change with age and others that do not. The study had the further advantages of relatively large sample size and availability of information on health status and medication of the subjects.

The Army Alpha has two parallel forms, A and B, each of which consists of eight subtests: Following Directions, Arithmetic Problems, Practical Judgment, Synonym-Antonym, Disarranged Sentences, Number Series Completion, Analogies, and General Information.

All subjects were given Form A of the test on their first visit. Thereafter, they were treated differently according to age. After an interval of four to eight years, those under 70 were given Form B, which is parallel to Form A, to avoid practice effects. Subjects aged 70+ continued to be administered Form A at each annual visit. This provided two sets of longitudinal data: retest data after four to eight years on Form B for 350 men initially aged 17 to 65 years, and data from six administrations of Form A spanning five to 11 years for 51 men initially aged 66 to 84 years.

The two-point data for men under 65 were extensively analyzed. In the first set of analyses, subjects were classified in three age groups, 20–39, 40–49, and 50–65 years, and were cross-classified as low, average, or high in blood pressure on the basis of three sources of data: average basal systolic and diastolic blood pressure at the first visit (designated Basal), average of casual sitting pressures recorded from right and left arms at the first visit (designated First Casual), and average right and left casual sitting pressures at the second visit (designated Second Casual). In the Basal classification, subjects were divided into low, middle, and high thirds of the distribution. In the First and Second Casual classifications, subjects were classed as "low" if their systolic pressure was below 120 or their diastolic pressure was below 80, "high" if their systolic pressure was above 140 or their diastolic pressure above 90. The remaining subjects were classified in a "middle" group.

In the first set of analyses, scores on each of the eight timed subtests and the total Alpha score were examined in relation to systolic and diastolic blood pressures under Basal, First Casual, and Second Casual classifications, a total of six definitions yielding up to 18 groups. Age group and time were also used as classifying variables. There was no evidence in the data for any effect of hypertension (within the existing blood-pressure ranges) on changes in Army-Alpha performance among men under 65. Similar analyses were conducted on data from the Army Alpha test given under untimed conditions. Although three of 48 main effects and five of 144 interaction terms were significant, the investigators concluded that this finding could not be considered evidence of any effect of blood-pressure level on changes in performance.

Since this was an unscreened population in which the influence of chronic illness or drugs could have obscured a true effect, further analysis was undertaken after exclusion of such individuals. Additional analyses were performed in which subjects taking medication or diagnosed as having a serious chronic illness, other than hypertension, at any point during their first five visits to the GRC were excluded. These exceptionally stringent criteria left a smaller sample of 117 men, few of whom were in the "high" hypertension group. Since there was no evidence of an interaction between age and hypertension, subjects were collapsed across age groups, and age was used as a covariate in the analysis.

When timed subtests and totals were treated as dependent variables, the "high"-blood-pressure group performed least well on the Following Directions and Number Series Completion subtests. Although the effects were weak and inconsistently replicated, they gave at least mixed evidence that higher blood pressure may be detrimental to some kinds of cognitive performance among otherwise healthy men. Somewhat stronger evidence was found when the untimed condition of the Alpha was analyzed. Replicated main effects were seen for Arithmetic Problems, Synonym-Antonym, Disarranged Sentences, and Number Series Completion, as well as for the total untimed score. The "high" group showed the poorest performance in all these measures. None of the effects, however, was strong, and no finding was statistically significant in all six possible replications. In these data no support could be found for the prevalent notion that elevated blood pressure or hypertension causes cognitive loss or decline with aging.

In an analysis of men over 70, all subjects were classified as a) free of drugs and major chronic diseases other than hypertension, or b) "other." Since no significant differences were found in any of the eight subtests or total scores, it was possible to collapse the two categories in subsequent analyses. Repeated-measures analyses of variance were performed on the Alpha subtests and total score, using six levels of the repeated factor and classifying subjects by age and blood pressure. On only one of the subtests, Practical Judgment, was there a significant main effect for blood-pressure classification, and then only for systolic, on which hypertensive men were somewhat lower performers than normotensive men. It might be expected, from the literature, that normotensives would remain relatively constant or show only a modest decline over the six administrations, while hypertensives would show a more progressive and precipitous decline. On the contrary, the study suggests that elevated blood-pressure levels have little if any effect on intellectual performance for men ranging from 17 to 88 years of age. The conclusion must be interpreted cautiously because of several limitations, which include the modest elevation of blood pressure and the absence of cases of severe hypertension.

PERSONALITY CHARACTERISTICS

1. Age and Stability of Personality

The question whether personality remains stable or changes over the adult life span has become a major focus of attention in the last decade, precisely the period during which the field of personality psychology was undergoing fundamental revaluation. Neugarten (1977) pointed out that the attacks of social learning theorists (Mischel, 1968), humanistic psychologists (Maddi, 1976), and psychometrists (Fiske, 1974) had left both personality psychology and its subfield, the study of personality and aging, in disarray. The dominant theories of adult development, outgrowths of psychoanalytic approaches, had not proven their usefulness. The critiques of Schaie (1965) and Baltes (1968), who demonstrated that cross-sectional studies confound age changes with birth-cohort differences, had brought into question most of the evidence on the relation of personality to aging.

It might have been expected that the resurgence of the field would require the development of a new conceptual model. Instead, the findings of stability argued for one of the oldest models of personality: trait theory. Previous trait theory (Eysenck, 1960; Cattell, 1973; Guilford et al., 1976), despite its discouraging theoretical and technical differences, had converged empirically on two basic dimensions of temperament, which may be labeled Neuroticism and Extraversion. A new trait model, the NEO schema (Fig. 1) (Costa and McCrae, 1980c), which added the dimension of Openness to Experience, provided support for earlier BLSA studies that had consistently found that adult personality remains stable with advancing age.

Statistical definitions of stability and change. The term "stability" has three different and largely independent meanings. A trait may be considered stable for a group if the mean level of the trait in a group of individuals is constant over time. This will occur if all individuals remain at the same level, or if equal numbers increase and decrease over the interval. Analysis of variance on age groups for repeated administrations is the usual way to determine whether significant changes in mean level have occurred. Another approach compares a person's standing on a trait in relation to that of others in the group. Stability of this kind is usually assessed by a test-retest coefficient, which will be higher if individuals maintain the same relative ordering on the trait over time, regardless of the level of the trait. If some developmental process leads to a uniform increase or decrease of a variable over time, it will have no effect on the retest coefficient.

The implication of these considerations is that, except in the artificial case in which all individuals score identically on repeated administrations of a test, the issues of mean-level stability and retest stability must be addressed separately.

There is a third way in which stability or change in personality might be seen across the life span. Personality is often conceptualized in terms of the relation between discrete variables. The pattern of intercorrelations among a group of traits might alter with maturation; and although this kind of change is least familiar, it is logically prior to the other kinds of stability or change. The most common method of comparing patterns of intercorrelations is factor analysis of the battery of tests showing that the same (or different) factors emerge in different age groups or administrations. "Age-comparative

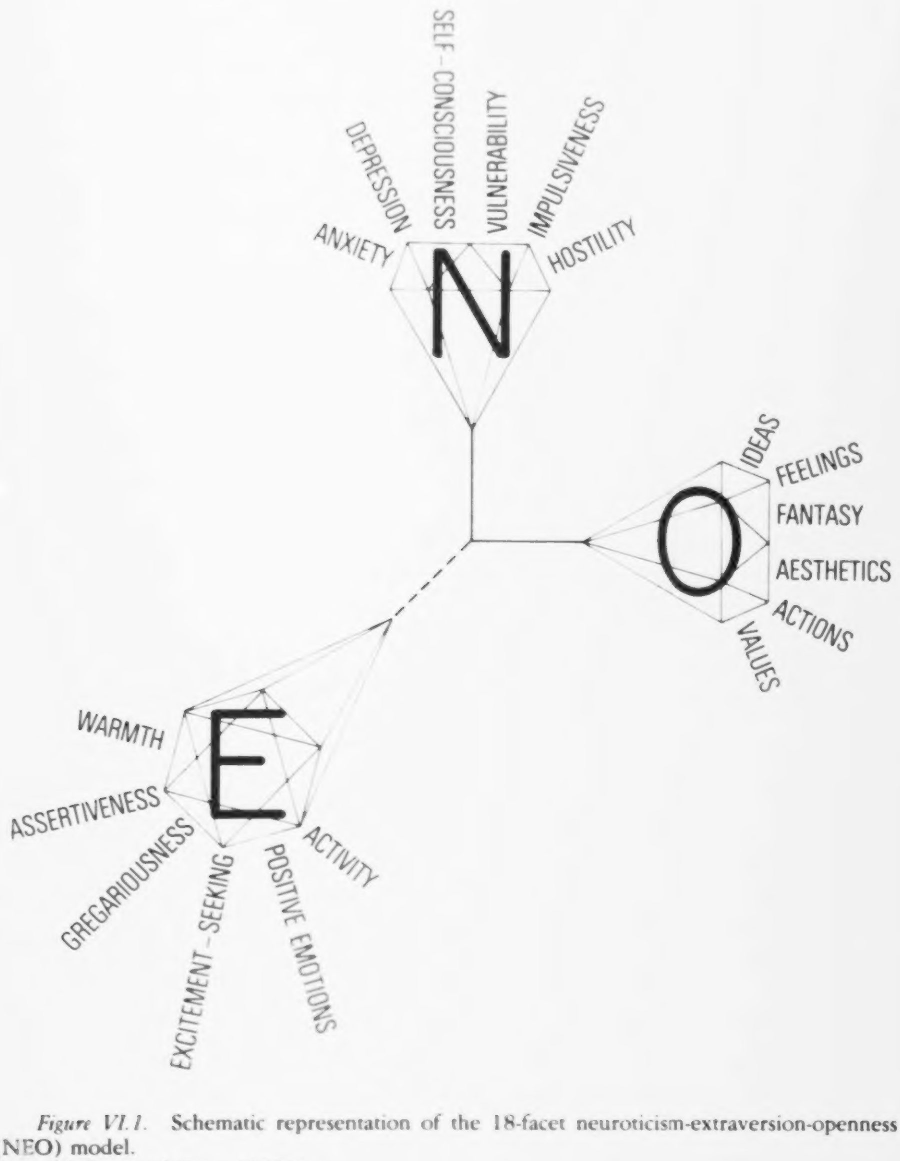

N
ANXIETY
DEPRESSION
SELF-CONSCIOUSNESS
VULNERABILITY
IMPULSIVENESS
HOSTILITY
O
IDEAS
FEELINGS
FANTASY
AESTHETICS
ACTIONS
VALUES
E
WARMTH
ASSERTIVENESS
GREGARIOUSNESS
EXCITEMENT-SEEKING
POSITIVE EMOTIONS
ACTIVITY

Figure VI.1. Schematic representation of the 18-facet neuroticism-extraversion-openness (NEO) model.
From McCrae and Costa (1980).

factor analysis" is a term occasionally used to designate this kind of analysis, while the type of stability is usually referred to as "structural stability."

The three definitions of stability were applied in a series of BLSA studies *(Douglas and Arenberg, 1978; Costa et al., 1980b; McCrae et al., 1980)* that investigated the question whether personality remains stable or changes with age.

Stability of group mean level. Evidence of the constancy of group mean level is provided by a study by *Douglas and Arenberg (1978),* which included cross-sequential and time-sequential designs in the analysis of a large sample on a standard personality instrument over a considerable span of time. The authors of the study investigated the relation between age and the ten scales of the GZTS in light of the following questions: a) Does the GZTS demonstrate age differences? b) Does it demonstrate age changes within individuals? c) Are age changes in repeated measures of the same subjects confirmed in cross-sequential analyses of the same birth cohort measured at different times? d) Are the changes found in repeated measures of the same individuals also found in time-sequential analyses (i.e., are the changes due to period effects rather than to aging)?

Between 1958 and 1974, the GZTS was administered to 915 BLSA participants from 17 to 98 years of age. Repeated measures were obtained for 336 men from 5.6 to 9.9 years after initial testing. Each GZTS scale was analyzed cross-sectionally and longitudinally. Time- and cross-sequential analyses of independent samples were also performed in an attempt to separate the effects of aging from those of birth cohort and period.

Only two scales (General Activity and Masculinity) were interpreted as showing aging effects: Preference for fast-paced activity started to decline at age 50; masculine interests declined at all ages. In general, results provide strong evidence for the stability view of adult personality. Only two of the ten scales showed evidence of age-related change, and the amount of change was very small for both. Personality differences among individuals at all ages are far more pronounced than the effect of age itself.

Stability of individual rank order. Retest correlations or stability coefficients assess the magnitude of personality consistency or change in the relative ordering of individuals, regardless of absolute level. These are among the most important analyses for longitudinal studies, for although different samples at different times, or different birth cohorts at one time, can be used to estimate age changes in trait levels, only repeated testing of the same individuals can speak to the degree of stability of individual differences.

The stability of individual rank order on the scales of the GZTS was examined longitudinally in BLSA participants who were tested on three occasions over a period of about 12 years *(Costa et al., 1980b).* Of primary interest was the degree of stability over such extended intervals. In addition, two subsidiary hypotheses were tested. The first hypothesis was that certain socially desirable traits, like sociability, assertiveness, and others that define the broad domain of Extraversion, are particularly stable, and that change is more characteristic of undesirable traits that might be interpreted as elements of the Neuroticism domain. The second hypothesis asserts that stability coefficients will increase with age. This hypothesis was based on evidence that stability coefficients increase for children (Nesselroade and Baltes, 1974), and suggestions that personality is increasingly consolidated in old age (Neugarten, 1964). The subjects were 460 BLSA participants ranging in age from 17 to 85 at the time of first testing. Subjects were readministered the GZTS at intervals of approximately six years over a

Table VI.1. Six- and 12-Year Retest Coefficients for GZTS Scales for Total Samples and 12-Year Retest Coefficients for Three Age Groups

	6-Yr Retest	12-Yr Retest	12-Yr Retest		
Scale	Total Sample (17–85 yr)[a]	Total Sample (20–76 yr)	Young Group (20–44 yr)	Middle Group (45–59 yr)	Old Group (60–76 yr)
General Activity	.83 (410)[b]	.77 (192)	.77 (60)	.82 (93)	.78 (39)
Restraint	.71 (418)	.72 (193)	.61 (62)	.74 (94)	.76 (37)
Ascendance	.82 (401)	.83 (194)	.85 (62)	.85 (95)	.77 (37)
Sociability	.81 (393)	.74 (182)	.64 (62)[c]	.81 (88)	.66 (32)
Emotional Stability	.74 (427)	.70 (203)	.63 (68)	.76 (96)	.71 (39)
Objectivity	.71 (405)	.69 (191)	.66 (64)	.76 (87)	.59 (40)
Friendliness	.77 (418)	.74 (193)	.74 (64)[d]	.68 (88)[e]	.87 (41)
Thoughtfulness	.72 (418)	.73 (199)	.78 (64)	.71 (94)	.71 (41)
Personal Relations	.73 (385)	.68 (188)	.70 (62)	.64 (89)	.73 (37)
Masculinity	.75 (417)	.72 (200)	.73 (66)	.71 (94)	.70 (40)
Mean stability	.77	.73	.72	.75	.73

[a] Numbers in parentheses in column headings are age ranges at Time I. All correlations are significant at $p < .001$.
[b] Ns are given in parentheses.
[c] Difference between young and middle groups is significant at $p < .05$.
[d] Difference between young and old groups is significant at $p < .05$.
[e] Difference between middle and old groups is significant at $p < .01$.
From Costa et al. (1980b)

12-year period. Three age groups, based on the date of first testing, were formed: young (17–44 yr, $\bar{x}$ age = 36.7 yr, 145 men); middle (45–59 yr, $\bar{x}$ age = 51.5 yr, 183 men); and old (60–85 yr, $\bar{x}$ age = 67.9, 132 men).

Table 1 gives the six-year and the 12-year stability coefficients for the total samples, and the 12-year stability coefficients within each of the three age groups.

Under the hypothesis that stability should be greater in older age groups, one-tailed tests of the significance of differences between correlations were computed for each pair of age groups on each scale. Six- and 12-year stability coefficients were exceptionally high, comparable to the short-term retest reliability. Coefficients were quite similar for the three age groups. Of the 30 comparisons of stability coefficients in different age groups at each interval, only six were significant for the six-year interval and three for the 12-year interval. Of the nine significant differences, four were in the predicted direction, five in the opposite direction. None of the specific findings for scales at the six-year interval was replicated at the 12-year interval. The hypothesis that younger men would show lower stability of personality traits than older men was not supported by the data. When corrected for unreliability, data were also inconsistent with the hypothesis of differential stability for different traits. These data provide strong evidence for the stability of personality traits in adulthood. Increased stability with age may be found among children and adolescents, but by young adulthood stability in these dimensions of temperament is so high—near the limits of reliability of the instrument—that a ceiling effect diminishes the likelihood of any further increase in stability. Personality stability across time and instruments provides support for one implication of the general stability position: that the time at which personality is measured is irrelevant. Once the individual has reached full adulthood, by age 30, a single measurement of personality would suffice for a lifetime. In other words, personality might be considered a constant throughout the individual's adult life span, including old age.

There are, of course, a number of reasons to stop short of so sweeping a stand. It assumes a perfect correlation of tests over time, a condition rarely observed even when corrections for unreliability are made. It also assumes that there are no circumstances under which personality might regularly be expected to change. Evidence to date shows only that the events encountered over the course of a lifetime by volunteer subjects do not systematically produce change. But other circumstances, such as therapeutic interventions, cataclysmic events, or severe illnesses, might affect traits. The longitudinal study of personality in adulthood is simply too young a field to rule out these possibilities.

On the other hand, if we turn from the individual to the group, much stronger arguments can be made. On the aggregate level, there appear to be good grounds for claiming that the time at which measurements are made should not affect the results; in particular, the relations between different tests ought not to depend on the times at which they are administered. Data collected in the 1960s might be used to validate tests created in the 1970s, just as if the administration had been contemporaneous.

This approach was adopted in a study relating archival GZTS scores to recently collected personality scores *(Costa and McCrae, in press)*. The criteria of primary interest were scales from the GZTS (Guilford et al., 1976), administered from 1959 to 1979 to subjects on their first or second visit to the GRC. Although the instrument was subsequently readministered, for purposes of simplicity only first-administration data are analyzed here.

Over the same period, data were also collected on the Cornell Medical Index (CMI). Twelve sections deal with physical symptoms, six with psychiatric complaints. Scores from the WAIS Vocabulary scale and the Army Alpha total collected in the same two-decade period were also used.

All these measures were used to consider the convergent and discriminant validity of three personality instruments administered in the past four years. Form A of the Eysenck Personality Inventory (EPI), with scales for Extraversion, Neuroticism, and Lie, was mailed to subjects to be completed at home in September 1979. The NEO Inventory was mailed to subjects in February 1980, and the NEO Rating Form was completed at home in August 1980 by spouses of subjects whose husbands or wives were also participants in the BLSA. Because some spouses did not participate, rating data are available for only a subset of subjects.

The NEO Inventory measures six facets, or aspects, of three global domains of personality. Neuroticism is represented by anxiety, hostility, depression, self-consciousness, impulsiveness, and vulnerability. Extraversion includes scales for warmth, gregariousness, assertiveness, activity, excitement-seeking, and positive emotions. Openness to Experience is measured in the areas of fantasy, aesthetics, feelings, actions, ideas, and values. Total scores for the three domains are obtained by summing the scores of the six facets in each. The analyses provided evidence of validity against new criteria; in addition, since the construct validity of the NEO scales is already fairly well established, they provided a demonstration of the feasibility of using data collected several years previously in validation studies.

The correlations between GZTS scales administered between 1966 and 1979 and

NEO facets administered in 1980, at least one year after the last GZTS, are presented in Table 2. The average interval between the two tests was 9.1 years. Despite this considerable lapse of time, construct validity is very much in evidence. General activity from the GZTS is most strongly correlated with NEO activity; GZTS ascendance with NEO assertiveness; GZTS sociability with NEO warmth; GZTS emotional stability and objectivity (negatively) with NEO anxiety; GZTS friendliness (negatively) with NEO hostility; GZTS thoughtfulness with NEO openness to ideas. At a more global level, the GZTS Extraversion scales, (G, A, and S) are correlated chiefly with NEO Extraversion scales; GZTS Emotional Health scales (E,O,F,P, and M) show consistent negative correlations with NEO Neuroticism scales. Finally, the magnitude of the correlations requires comment: Almost all exceed .30, and several reach .60. In view of the somewhat different conceptions underlying the two instruments, these are high correlations.

Correlations with NEO ratings by spouses are generally smaller than correlations with self-reports and occasionally fail to reach statistical significance, in part because they are based on considerably fewer cases. But in view of the fact that the correlations cross instruments, methods of measurement, and two to 20 years, the pattern of results is convincing. A correlation of .59 between GZTS Extraversion and spouse-rated NEO Extraversion is particularly remarkable.

2. Historical Effects on Test Data

As the value of longitudinal studies becomes more apparent, we can expect more researchers to collect and archive personality data, or to retrieve and retest subject populations measured some time before. Longitudinal methodologists have pointed out that there are potential problems in the use of such data, since the meaning of the test may have changed in the intervening years. This phenomenon can readily be envisioned in the case of attitudinal research: Attitudes toward women that were radical in the 1950s might be middle-of-the-road today.

In the case of personality measures, this concern appears to be less well founded. Over the past 20 years, scales measuring the major dimensions of personality do not appear to have altered substantially in meaning. Subjects today appear to respond to test items in much the same way as they did in the 1950s and 1960s. Basic aspects of temperament are relatively impervious to historical changes, and longitudinal research can use older personality data with some confidence that the meaning of test items has not changed.

Indeed, it seems more likely that what has changed is our conceptualization of what older tests measure. For example, the Bernreuter (1933) inventory contained a scale labeled "introversion" that we might now conceptualize instead as social anxiety, a part of Neuroticism. Such hypotheses are testable if we readminister older instruments and revalidate them against measures that are better conceptualized and psychometrically more sophisticated. Since the relations between tests appear to change little over an interval of years, we could use the results of contemporary studies to reinterpret older findings on the basis of our current understanding of the nature of the constructs they measure.

Stability of factor structure. Constancy and change in adult personality organization can also be assessed by comparison at different ages of the factor structures of personality instruments. This approach was used in another BLSA study by *McCrae et al. (1980),* which compared the factor structure of the GZTS longitudinally in three

Table VI.2. Correlations between GZTS Scales Administered 1959-1966 and NEO Facets Administered in 1980

	GZTS Scales[a]				
NEO Facets	G	R	A	S	E
Neuroticism:					
Anxiety	-14	-01	-25**	-25***	-67***
Hostility	23**	-24**	07	-12	-47***
Depression	-04	-02	-38***	-38***	-58***
Self-Consciousness	-12	-04	-37***	-33***	-55***
Impulsiveness	12	-35***	16	11	-44***
Vulnerability	-31***	08	-44***	-34***	-39***
Extraversion:					
Warmth	11	-08	31***	56***	18*
Gregariousness	31***	-25***	41***	52***	17
Assertiveness	34***	-14	61***	41***	20*
Activity	61***	-18*	20*	10	03
Excitement-Seeking	41***	-34***	42***	21***	-13
Positive Emotions	32***	-29***	27**	30***	10
Openness:					
Fantasy	07	-12	10	-06	-30***
Aesthetics	19	10	26**	26**	19*
Feelings	22*	00	31***	21**	-09
Actions	22**	-10	31***	20*	03
Ideas	12	13	29***	08	07
Values	10	-12	23**	-05	-14
	O	F	T	P	M
Neuroticism:					
Anxiety	-55***	-31***	17	-31**	-28***
Hostility	-44***	-53***	04	-26**	-24**
Depression	-52***	-26**	11	-29***	-26**
Self-Consciousness	-46***	-25**	02	-22**	-26**
Impulsiveness	-31***	-46***	-06	-27**	-13
Vulnerability	-38***	-03	01	-10	-26***
Extraversion					
Warmth	15	08	10	12	-03
Gregariousness	16	-05	-23**	16	00
Assertiveness	15	-20	01	01	00
Activity	01	-16	01	-08	06
Excitement-Seeking	-08	-33***	-05	01	02
Positive Emotions	07	-06	10	-07	-05
Openness:					
Fantasy	-20*	-22**	22**	-22**	-02
Aesthetics	15	13	31***	06	00
Feelings	-06	-24**	31***	-12	-15
Actions	05	00	09	08	24**
Ideas	08	05	35***	09	11
Values	-13	-11	19*	-11	00

[a] GZTS scales are General Activity (G), Restraint (R), Ascendance (A), Sociability (S), Emotional Stability (E), Objectivity (O), Friendliness (F), Thoughtfulness (T), Personal Relations (P), and Masculinity (M). N = 140 to 152. Decimal points omitted. *p < .05; **p < .01; ***p < .001.

From Costa and McCrae (in press)

administrations, three age groups, and two times of measurement. Factor analyses were performed on correlation matrices computed for eight groups. Longitudinal comparisons were based on analyses of first-administration data from 769 men, second-administration data from a subset of the first group consisting of 346 men, and third-administration data from a further subset of 171 men. Cross-sectional comparisons were based on first-administration data from three age groups: young (17–44 yr, $\bar{x}$ age = 34.4 yr, 314 men); middle (45–59 yr, $\bar{x}$ age = 51.6 yr, 242 men); and old (60–97 yr, $\bar{x}$ age = 70.4 yr, 213 men). Possible structural differences resulting from period effects were assessed by dividing the sample into two groups: 455 men who first took the GZTS before July 1968 (age range 17–83 yr, $\bar{x}$ age = 52.1 yr), and 314 men who first took the GZTS after that date (age range 18–96 yr, $\bar{x}$ age = 45.6 yr); these analyses were limited to data obtained at the first administration of the test. All factor analyses were restricted to subjects with valid scores on all ten GZTS scales.

Factor structures in the GZTS in this sample were clearly invariant. Despite aging, attrition, and possible practice effects, the same pattern was seen at each administration. The data revealed no meaningful differences when GZTS measurements before and after 1968 were compared, despite great social differences between the 1960s and 1970s; the implication is that the basic structure of personality is not greatly influenced by social and historical change in the life of the individual.

3. Eliminating Response Bias as an Explanation for Stability

Response sets and age. Most research on the stability of personality has been conducted with self-report measures, which have in general shown a better record of internal consistency, retest reliability, and construct validity than ratings or projective tests.

But self-report inventories are also prone to certain problems. The transparency of the items makes it possible for individuals so motivated to present themselves favorably or unfavorably. The use of a standard format for answering questions (yes-no, Likert scale, a rating bar) makes it possible for consistencies in the style of response to distort the scores obtained from the instrument; the result may be spuriously high consistency or correlations. The effects of social desirability, acquiescence, and extreme response have been the source of interminable debate among personality psychologists, and no definitive resolution has been reached.

In addition, there has been relatively little research on age trends in response sets. Some writers (Schaie and Schaie, 1977) have argued persuasively that there have been enormous changes in the amounts and kinds of testing to which individuals of different generations have been exposed, and that this may introduce unwanted sources of variance in tests. Increased cautiousness may alter the responses of older subjects (Botwinick, 1969), or standards of social desirability may change with age, bringing shifts in the influences of that set. Age changes or cohort differences in response could account for observed differences in scale scores or could mask real changes that are occurring. To date, these possibilities remain speculative, with little empirical foundation. Clearly, before any conclusions about aging and personality are drawn from objective personality measures, some information on these issues would be useful. Recent analyses of data from the Baltimore study *(Costa et al., 1983)* contribute to a resolution.

Over the past 20 years, subjects in the BLSA have been given the GZTS every six years. Since recruitment into the study has been continuous, new samples of individuals, ranging in age from the 20s to the 90s, have been tested successively. If their ages, birth cohorts, and times of measurement are divided into six-year intervals, a variety of analytic designs may be applied to aid in the interpretation of changes or differences.

The GZTS was given to all subjects with standard instructions that provide three response options—"yes," "no," and "?"— but subjects were instructed to use the "?" option only if they were completely unable to select "yes" or "no." In conformance with the suggestion of Guilford and Zimmerman (1949), scales containing more than three "?" responses were invalidated in all previously reported applications of the GZTS conducted by the Baltimore study. This exclusionary principle may, however, distort results; in particular, if age produces caution in responding, a disproportionate number of older subjects may be excluded, perhaps especially the most cautious. Because of that possibility, a new approach was adopted in the analyses reported here. All the GZTS answer sheets were keypunched, so that responses to individual items could be analyzed. The original scoring system, in which "?" responses were not scored, tended to lower the score of the individual. In the new system, responses were assigned a value of -1, 0, or $+1$, the "?" being represented as a neutral, rather than negative, value.

The handbook for the GZTS (Guilford et al., 1976) lists three scales that have been developed to estimate the influence of certain response sets: the Gross Falsification (GF), the Subtle Falsification (SF), and the Careless Deviancy (CD) scales. The first two are intended to screen individuals who may be attempting to present an unduly favorable impression; the third is composed of relatively rare responses, a high score on which is interpretable either as careless responding or as deviancy in personality. In addition, it is possible to measure at least three other response sets on the GZTS. The number of blanks was summed across all scales to give an index, as was the number of "?" responses. Since items in the GZTS are roughly balanced on most scales, it is possible to interpret the sum of "yes" responses as an index, not of any substantive personality trait, but of the tendency to acquiesce indiscriminately.

For each of these variables, cross-sectional, longitudinal, cross-sequential, and time-sequential analyses were performed. In the repeated-measures analysis, the subjects were 348 men ranging in initial age from 32 to 74 who were retested after four to eight years ($\bar{x}$ interval = 6.6 yr). They were classified in seven age groups, each spanning a six-year interval. In the time-sequential analyses, 328 men who were tested in the period from 1958 to 1964 were compared with 278 men tested between 1965 and 1971. They were cross-classified by the same seven age categories used in the repeated-measures design. In the cross-sequential analysis, 345 men tested between 1958 and 1964 were compared with 285 men tested between 1965 and 1971. In this design, however, they were cross-classified in seven cohorts born during six-year intervals from 1896 to 1932.

Repeated-measures analyses showed significant ($p < .05$) effect on the repeated factor for number of question marks, which increased from 6.2 to 9.6; on acquiescence, which decreased from 132.1 to 128.9; on the GF scale, which increased from 11.8 to 12.1; and on the SF scale, which decreased from 21.2 to 20.8. Age-group differences were seen for the number of question marks, which was highest in the 68–74-year-old group and lowest in the 50–55-year-old group. In addition, there were two interactions: Men aged 38 to 43 at first testing showed a decrease instead of an increase in question marks, and men aged 55 to 61 showed an increase instead of a decrease in acquiescence.

These results are somewhat puzzling; certainly they did not show a monotonic change in any of the response sets with age. Most of the changes were extremely small in magnitude, and if we require a significance level of $p < .01$, only two effects are significant: the increase in question marks and the decrease in acquiescence. The fact that neither of these longitudinal changes was mirrored in cross-sectional differences suggests that the changes are due either to time-of-measurement effects (a cultural change during the testing period) or to "practice" (repeated exposure to the test).

Examination of the cross- and time-sequential analyses, conducted on samples of more than 600 men, is revealing. Analyses of number of blanks, number of question marks, acquiescence, GF, SF, and careless deviancy show no significant ($p < .05$) effects for aging/cohort, aging/time, cohort/time, or cohort/aging, nor were there any significant interactions. These data suggest that the marginal cross-sectional differences and interactions in the repeated-measures analyses are best regarded as unreplicable error and that the longitudinal changes in acquiescence and use of question marks are attributable to practice effects. That acquiescence decreased while use of question marks increased by three items is suggestive: Subjects may have felt pressured on the first administration to avoid question marks at all costs and may thus have agreed to a few items of which they were uncertain. Some years later, as experienced subjects no longer so hesitant to assert themselves, they may have used the question marks when they felt they needed to.

In any case, these data imply that response sets are not ordered by age. Although longitudinal research may want to consider the effects of repeated administration of the same instrument, the particular effect seen here is small in its overall influence on scale scores and is probably unique to instruments, like the GZTS, that provide a question-mark option but fail to score it.

4. The Stability Model: Implications for Gerontology and Geriatrics

These results favor the stability model for objectively measured personality traits throughout adulthood, at least for males. The implications for gerontology and geriatric practice of these consistent findings are radical. Neugarten's classic formulation that adult development is "the changing basis within an individual for adaptation to life" may well have to be changed to read "the *stable* basis." Students of the relation of personality to aging, seeking the basis of the impact of major life events, may well have to shift the focus of their efforts from the vectors of personality change to the mechanisms by which personality preserves equilibrium. Not least significant is the corollary that a clinician who finds true personality change in an aged patient is likely dealing not with a normal event but with evidence of disease that can often be treated.

These findings do not prove that personality is unchangeable. Some individuals do change in one or more characteristics, for reasons not yet understood. It is reasonable to suppose that psychotherapeutic intervention can make real changes in personality, and a host of techniques, from cognitive behavior modification to biochemical intervention, may have profound effects as yet undocumented. What we can say is that such changes, for better or worse, are not likely to happen to anyone simply as a result of growing older.

5. Longitudinal Changes in Adjustment to Aging

Adjustment to aging and the effect of age on personal adjustment among the aged have long been of concern to gerontologists. A recent study (*Costa and McCrae, 1982*) used cross-sequential and time-sequential as well as traditional longitudinal designs to examine the stability or change in the Chicago Attitude Inventory (CAI). The eight sections of the CAI—Health, Friends, Work, Economic Security, Religion, Usefulness, Happiness, and Family—were supplemented by two global items concerning assessment of life happiness and satisfaction with accomplishments in life. Analyses were conducted on a sample of 425 men, aged 17 to 97, on whom first-administration CAI data had been collected before 1970. Results of cross-sequential, time-sequential, and repeated-measures analyses led to the conclusion that attitudes toward usefulness and work showed small age-related declines that were interpreted as non-maturational. Assessment of Life, Satisfaction with Accomplishments, and attitudes toward Health, Friends, Economic Security, Religion, Happiness, and Family showed no consistent changes with age. Personal adjustment was thus found to be quite stable across the adult and later years.

6. Personality and Adjustment to Aging

Enduring dispositions as predictors of successful personal adjustment to aging were examined in a BLSA longitudinal study by *Costa et al. (1981)*. Successful aging can be studied by focusing a) on outer social adjustment defined in terms of an individual's activities and social roles, b) on the inner subjective experience of personal adjustment, or c) on subjective well-being. From the second perspective, satisfaction with life's accomplishments, retention of high morale, or simple happiness is considered evidence of successful aging, while hypochondriasis, fear of death, and a sense of uselessness, loneliness, and depression are considered signs of poor adjustment to aging.

The second definition was used in this BLSA study, which examined the ability of longitudinal measures of Neuroticism and Extraversion to predict successful personal adjustment to aging. The study used the GZTS to measure personality characteristics and the CAI to measure subjective well-being. The ten GZTS scales yield scores on three factors: Neuroticism, Extraversion, and Thinking Introversion (a factor thought to represent meditative thinking or introspective tendencies). It was hypothesized that Neuroticism would have a negative relation to CAI measures of subjective well-being, that Extraversion would have a positive relation, and that Thinking Introversion would have no relation to those measures.

The CAI and the GZTS were administered to each of the men in the sample (aged 17–97 yr) on the first or second visit to the GRC. The CAI was readministered on the fifth and ninth visits. In the first analyses, correlations between the CAI scales and the contemporaneous GZTS factors were calculated for two subsamples: 418 men aged 18 to 49 and 391 men aged 50 to 97 years. In the second analysis predictive relations between personality and adjustment were sought by correlation of GZTS Neuroticism and Extraversion scores from the first administration with scales from the second and third administration of the CAI for the 577 men with repeat administration data.

As hypothesized, Neuroticism was related negatively and Extraversion positively to most concurrent measures of well-being in both younger and older subsamples, and Thinking Introversion was unrelated to the well-being measures. Predictive correlations between personality and subjective well-being over intervals of two to ten years ($\bar{x}$ interval = 5.3 yr) and ten to 17 years ($\bar{x}$ interval = 12.6 yr) confirmed earlier research (Costa and McCrae, 1980b) and showed that enduring personality dispositions precede and predict measures of personal adjustment to aging.

Using more sophisticated measures of well-being, *Costa and McCrae (in press)* have shown that an individual's affect balance and life satisfaction can be predicted years in advance by assessment of personality. They used GZTS scores collected in the period from 1959 to 1969 with well-being data collected ten to 23 years later ($\bar{x}$ predictive interval = 15.6 yr for first and 17.7 yr for second interval). As had been predicted by a model of psychological well-being (Costa and McCrae, 1980b), three of the four Extraversion scales were significantly related to well-being, especially to positive affect, while four of the five Neuroticism scales were significantly related to well-being, especially to negative affect. These correlations, which were not markedly different in size from contemporaneous measures, provide a strong longitudinal replication of the model.

7. Somatic Complaints as a Function of Age and Neuroticism

Self-perceptions of health are key components of health maintenance, since they influence efforts at self-medication as well as decisions to seek medical treatment. Studies have consistently shown that global self-ratings, which are moderately correlated with medical determinations of health, are also related to such psychological characteristics as health attitudes, morale, adjustment, and psychological distress. Despite the demonstrable increase in many kinds of illness with age, global self-ratings often fail to show any marked association with age.

Symptom checklists like the CMI offer certain advantages over global ratings. Since they ask specific questions about conditions and symptoms, they may be less influenced by general health attitudes and can be used to analyze medical conditions of body systems separately, as well as to determine whether age is differentially associated with complaints in particular somatic systems. The CMI is a self-report symptom checklist with 195 items divided into 12 somatic sections (A–L) and six psychiatric sections (M–R). Somatic sections I (Frequency of Illness) and J (Fatigue) had extremely low endorsement in this sample and were therefore combined in all analyses. The sum of sections A–L yields a measure of total physical complaints, while the sum of sections M–R provides a measure of psychiatric complaints. Like global ratings, checklists are influenced by both objective health and psychological factors. Hypochondriasis, neurosis, general anxiety, poor marital adjustment, and psychological problems have all been found to be linked to extremely high endorsement of CMI physical complaints.

Many of the psychological factors identified in previous research can be hypothesized to be related to Neuroticism. The hypothesis was tested in a BLSA longitudinal study that examined the relative influences of age and Neuroticism on self-perception of health or illness and on total physical complaints *(Costa and McCrae, 1980a)*.

At the time of their first administration of the CMI, the 1038 subjects ranged in age from 20 to 97 years. Although some analyses used data from a second or third administration of the CMI, the number of subjects was smaller because of insufficient length of participation in the study, death, or withdrawal. Endorsement of psychiatric items was low in this population, about 40% of the subjects endorsing none and 20% only one of the 51 items. The Emotional Stability Scale of the GZTS was used as an alternate measure of neuroticism. This is a 30-item scale covering Evenness vs. Fluctuation of Mood; Perseveration of Ideas; Composure vs. Excitability; Daydreaming; Feelings of Guilt, Loneliness, or Worry; and Cheerfulness vs. Gloominess. Two items explicitly refer to feelings of Good vs. Ill Health. Correlation between the GZTS Emotional Stability Scale and the CMI psychiatric score was found to be 0.52 in a sample of 915 subjects.

Longitudinal analyses were restricted to subjects whose second and third administrations of the CMI were, respectively, five to eight years and ten to 17 years later than the first. Each subject had also completed the GZTS Emotional Stability Scale at either the first or second visit to the GRC. Four sets of repeated-measures analysis of variance were conducted, with age and Neuroticism as classifying variables. In all analyses, subjects (N = 248) were classified in three groups of equal size as young (20–44 yr), middle-aged (45–56 yr), or old (57+ yr) on the basis of their ages at the first administration. In two of the sets of analyses, Neuroticism was measured by the GZTS Emotional Stability Scale and subjects were classified as unstable or stable. In the other two sets, Neuroticism was measured by the CMI psychiatric (M–R) score, and subjects were classified as high or low (subjects endorsing only one item in the M–R section were omitted from these analyses).

Supplementary cross-sequential analyses using different subsets of the same cohort (551 subjects), and time-sequential analyses (637 subjects) were also conducted on data collected at the first administration. The data showed a consistent pattern across different methods of analysis and different measures of Neuroticism. Age had a selective effect on physical complaints, while Neuroticism appeared to produce a more general and diffuse effect. Problems in sensory, cardiovascular, musculoskeletal, and genito-urinary systems increased with age, while health habits improved. The fact that only certain systems showed age-related increases in complaints may in part account for the finding that global health ratings are only weakly related to age. The finding of significant effects of Neuroticism despite its limited range in this healthy and well-adjusted BLSA sample argues that the relation between Neuroticism and physical complaints must be quite strong. The influence of Neuroticism on health perception may be even more pronounced in the general population.

It was concluded that research in the influence of physical health on morale, sick-role behavior, or adjustment should either use objective measures of health or supplement self-ratings with measures of Neuroticism in order to control for its effects. Most important, the findings convincingly contradict the conception that the aged are typically hypochondriacs obsessed with their bodily functions.

8. Hypertension, Somatic Complaints, CAD, and Personality

Hypertension, Hypochondriasis, and Neuroticism. Reports that hypertensives score higher than normotensives on measures of maladjustment, hypochondriasis, and neuroticism have led some writers to conclude that hypertension is related to personality (Sainsbury, 1960). Specifically, the chronic internal stress of repressed rage or anger is thought to lead to the development of hypertension. However, two alternative hypotheses are also consistent with the data. First, the association may be artefactual, the result of self-selection among clinical populations. Under this hypothesis, neuroticism leads only to the discovery or detection of hypertension. Second, the causal relation may be reversed; hypertension itself, the patient's awareness of illness, or medical treatment may lead to poorer psychological adjustment.

In an attempt to decide between these alternatives, *Costa et al. (1980a)* examined longitudinal data on systolic and diastolic blood pressure and data from the CMI and

GZTS from more than 700 BLSA subjects. Of the volunteers, 101 who were taking medication for hypertension either when they entered the study or at the succeeding four visits were eliminated from all analyses because of the possibility that their medication might increase somatic concern while decreasing blood pressure, and thereby obscure any real positive association between the two variables.

Each subject was given the GZTS and CMI during his first or second visit to the GRC. The GZTS was readministered approximately every six years, the CMI on the fifth visit. Longitudinal analyses were limited to subjects who took their second GZTS and second CMI four to 12 years after their first medical examinations. The average predictive intervals were 7.4 years for the GZTS and 6.5 years for the CMI. A total of 12 personality and physical-complaint variables were examined.

In cross-sectional analyses, higher somatic complaints, Restraint, Friendliness, and Good Personal Relations, as well as lower General Activity and Masculinity, showed small associations with higher blood pressure. However, when the effect of age (which is related to blood pressure) was partialled out, no statistically significant association between personality characteristics and blood pressure was found.

Does hypertension affect personality? If elevated blood pressure itself affects personality, the effect should have been evident in the correlations discussed above. It is possible, however, that the interval between a blood-pressure increase and the measurement of personality was not long enough for the hypertension to influence personality. To test this possibility, predictive analyses using average pressures at first measurement were correlated with GZTS and CMI scores collected four to 12 years later. Systolic pressure was correlated with greater Ascendance and Sociability and poorer Personal Relations, but the correlations were only marginally significant.

It can be argued that the GZTS, which is based on self-reporting, is inherently incapable of accurately detecting repressed rage and that clinical judgment is needed. An attempt was therefore made to approximate clinical judgment by examining combinations of personality scales. By cross-classification on scales F (Friendly vs. Hostile), R (Restrained vs. Impulsive), A (Ascendant vs. Submissive), and M (Masculine vs. Feminine), six hypotheses were tested: that higher average pressure might be found in individuals who were a) hostile but restrained, b) ascendant but restrained, c) masculine but restrained, d) hostile but submissive, e) masculine but submissive, or f) hostile but feminine. Neither main effects nor interactions proved significant when age was used as a covariate.

In summary, when the influence of age was controlled, no association was found between blood pressure and any of 12 personality and psychosomatic concern measures; nor did patterned combinations of traits show any relation to blood pressure. Attempts to predict personality scores four to 12 years later from blood-pressure levels also failed; the inference is that any changes in anxiety or somatic concern due to hypertension or its treatment are short-lived.

Does coronary disease affect personality? Examination of the relations between CAD and personality revealed that BLSA subjects who complained of angina were lower than average on the GZTS scales Emotional Stability and Masculinity. Was this an example of personality change resulting from medical illness, or were personality variables diagnostic of and perhaps causally involved in the development of the disease? To test out these possibilities, it was necessary to separate subjects into different groups and to take into consideration the temporal course of the disease *(Costa et al., 1982)*. Eighty-eight subjects were selected who were measured on the GZTS during their first or second visit to the GRC, and were free from both anginal complaints and certain ECG signs on these visits. Over the next twenty years, four groups emerged: those who developed CAD as evidenced by both anginal complaints and ECG signs; those who showed ECG signs of CAD, but did not report angina in a follow-up period of from five to 15 years; those who reported angina but showed no ECG signs of CAD in the same follow-up period; and a control group (age-matched to the total of the three groups) that showed neither ECG signs nor angina in a follow-up period of ten to 20 years. (All subjects were classified according to their status at last examination; the variation in the follow-up intervals reflects the fact that subjects entered and left the study at different times.)

The first group is easily diagnosed as having CAD. The second and third groups are more ambiguous: Those with only angina seem to be overly sensitive; those with only ECG signs appear under-sensitive. Do any of these distinctions show up in personality measures taken before the development of disease?

Results showed that there were no differences between the first and fourth groups, that is, between those who definitely did and did not develop CAD. Thus none of the traits measured by the GZTS appears to have etiological significance. But there were pre-existing differences between the two intermediate groups. Individuals who complained of anginal pains, but who gave no ECG evidence of disease, were less emotionally stable than those who reported no angina despite ischemic ECG signs.

It is well known (Hurst et al., 1978; Froelicher, 1977) that resting and stress ECG signs in themselves are far-from-perfect indicators of CAD, and the number of subjects in this study (88) is too small to be conclusive. Nevertheless, some interesting interpretations are suggested by the data. Men high in Emotional Stability may be more likely to minimize minor chest pains, so that without routine medical examinations those who have CAD are likely to remain undiagnosed and untreated (cf. Berglund et al., 1975). On the other hand, men who are low in Emotional Stability are very sensitive to chest pains, may report anginal symptoms even in the absence of organic pathology, and may request medical attention that is not required. Taken together, these analyses suggest that although the personality variables measured by the GZTS do not affect the development of CAD they seem to affect the presentation of symptoms.

If these personality differences were found at the same time as the symptoms, several alternate interpretations could be offered. We might argue that the experience of angina was sufficiently traumatic to lower the individual's emotional stability or we might argue that self-selection played the crucial role: Those individuals who were aware of chest pains and who were also predisposed to worry about their health (i.e., the more neurotic) would be most likely to join and remain in a longitudinal study that promised periodic monitoring of their health. Better-adjusted men with angina, some of whom could be expected to show no ECG signs and would thus be classified in the third group, would be less concerned with their health and less likely to volunteer for the study. These arguments have been supported in studies of the relationship between hypertension and personality *(Costa et al., 1980a)*.

In this case, however, we can rule out those alternatives. Since measurement of personality preceded the development of angina and ECG signs, it could not have been influenced by them. Self-selection on this basis is likewise impossible, since the individuals were presumably unaware that they would experience angina in the next few years. Retrospective accounts of personality difference could not be trusted in this context; only the archives of a longitudinal study permit the kinds of inferences drawn here.

These rather simple examples of the kinds of inferences that can be ruled out with longitudinal data are useful in part because they illustrate the underlying logic. Path-analytic techniques, however, offer far more sophisticated statistical models, which estimate not only the direction of causal influence but its degree, expressed as a regression weight (see Kenny, 1979). Causal inferences based on these statistical techniques are subject to a number of restrictions, many of them related to the assumptions necessary to construction of the model. In general, fewer assumptions are necessary and more reliable results can be obtained if more measurement points are used and more of the variables that might plausibly influence elements in the model are measured. Clearly, longitudinal studies are ideal for this type of analysis.

9. Research on Personality and the Life Course

The major drawback to the use of a stable individual-difference approach to life-span development is that current conceptions and methods of research on aging and personality are designed primarily for the study of change. The elegant models intended to separate true maturational changes from cohort differences and cultural changes are not very useful if there is no meaningful maturational change. The widespread attempt to chart the developmental course of personality in adulthood no longer seems profitable, and it may not be immediately clear what direction future research should take.

We have argued elsewhere (*McCrae and Costa, 1982*) that personality dimensions may be more usefully construed as causes than as effects. Age, and its attendant social, cognitive, and biological changes, should be considered in conjunction with personality as joint shapers of the life course. Personality can help explain the choices (educational, career, familial) that must be made at specific life transitions. The stability of personality dispositions may contribute to the individual's sense of identity and to the continuity and coherence of the life course. Finally, adaptation to life at all ages is likely to be powerfully influenced by personality.

Prospective life histories collected by longitudinal studies can help to answer such questions as the following: How do the lives of introverts differ from the lives of extraverts? Does openness to experience lead to a more fluid and unpredictable life course? How is neuroticism typically manifest at each stage of life? Answers to these questions may help integrate the insights of life-span developmental psychology with those of personality in the study of lives.

OUTCOME STUDIES

1. Obesity and Longevity

"Desirable" or "ideal" weight goals for adult men and women have been recommended since the turn of the century by the insurance industry on the basis of its experience with "insured lives." The tables have been and continue to be used widely—by the lay public, physicians, public-health officials, and clinical investigators. The most recent tables (Metropolitan Life Insurance Co., 1959) are based upon data subsequently published in a report of the Society of Actuaries (1960).

Since that publication, however, there have been reports of the obesity-mortality association in more than 40 other diverse populations (Andres, 1980a, 1981), most of them in the United States and western Europe with isolated reports from Japan, Australia, and Israel. There is a strong consensus that minimal mortality does not occur in the leanest segments of the population (as is commonly averred), but in individuals who range from the middle of the "desirable" weight range in the 1959 insurance tables to a value at least 20% over the mid-point. Indeed, one of the recent reports is an update of the experience of the insurance industry (Society of Actuaries and Association of Life Insurance Medical Directors of America, 1979); its data indicate that updated "desirable" weight tables will be required and that weights in them will be considerably increased over those of the 1959 tables.

Experience in the BLSA also indicates that mild or moderate overweight has been over-emphasized as a risk factor for mortality (Andres, 1980b). One of the possible confounding variables in the obesity-mortality link is cigarette smoking. Since smokers as a group weigh less than nonsmokers, it is theoretically possible that what appears to be the benefit of mild overweight may, in fact, be due to the fact that the percentage of smokers in that group is lower than that in the lean segment of the population. This hypothesis was tested in the BLSA subjects and was rejected *(D. Elahi et al., in press)*. In both current smokers and nonsmokers, there was a U-shaped relationship between obesity and mortality; mortality was highest in the leanest subjects in both groups, reached a nadir at moderately obese levels, and increased again in the more severely obese subjects. It must be emphasized that the BLSA population includes very few severely obese subjects. There is no question of the deleterious effects of severe obesity on longevity as well as on many specific diseases. Further analyses of the BLSA data will examine the impact of mild and moderate obesity on such end-points as hypertension, diabetes, and CAD.

2. Age and Some Physiologic Variables

A study by *Tobin (1981)* examined four physiologic variables clinically related to health. The variables included tests of the respiratory system (forced expiratory volume in one second), cardiovascular system (systolic blood pressure), renal system (standard creatinine clearance), and metabolism (glucose tolerance). Each is influenced by age and, at its extreme, is also indicative of a disease (chronic obstructive pulmonary disease, renal failure, hypertension, and diabetes). Normative data were obtained from individuals who were free of disease, poor health habits such as smoking or inactivity, or medications known to influence the organ system under study. This was done to ascertain the effect of age as such, in distinction from disease or other factors. To judge the importance of each variable, the age-adjusted T-scores for individuals who had died were compared with those individuals who had survived a ten-year interval.

There was no effect of age on the fasting plasma-glucose level. At two hours of a standard oral glucose tolerance test, there was a clear effect of age on blood-glucose concentration, with higher levels in each successive age group, although there was large variance at every age. Results of the tests of forced expiratory volume, systolic blood pressure, and creatinine clearance likewise showed a progressive impairment of performance across the age span, subjects in each successive age decade performing less well than the younger ones.

T-scores calculated within age decades for each of the 162 volunteers who had died since the BLSA was initiated were compared with the T-scores of those who were still alive (matched for age). T-scores for those variables whose values increased with

Table VI.3. T-Scores for Live and Dead Groups

		BP	$FEV_{1.0}$	C_{cr}	Glucose
Live					
	Mean	50.2	49.4	49.6	47.8
	SEM	0.33	0.37	0.35	0.46
	N	860	813	772	676
Dead					
	Mean	45.9	44.9	46.8	46.9
	SEM	0.77	1.08	1.01	1.20
	N	162	135	155	100
P		< 0.001	< 0.001	< 0.01	NS

From *Tobin (1981)*

age (blood pressure and glucose) reflected poorer performance. Measurements of blood pressure, forced expiratory volume, and creatinine clearance of those who had died had all been significantly poorer at the time of first testing than the scores of those who lived, and represented on the average a poorer age-adjusted performance for these values (Tab. 3). There was no significant difference between the two groups in glucose concentration at two hours after oral ingestion of glucose, although both were significantly lower than the normative levels.

Most of the deaths in this population have, as was expected, resulted from cardiovascular disease and cancer. As the study proceeds, more specific mortality data will be available to allow comparison of age changes in function commonly occurring with disease entities.

3. Predicting Coronary Events in Asymptomatic Subjects

Despite the much-publicized decline in mortality from CAD over the last decade, it remains the most common cause of death in the United States, especially in the elderly. Although the presence of standard risk factors (smoking, hypertension, hypercholesterolemia) or an abnormal ECG response to exercise is associated with an enhanced risk of future coronary events (angina pectoris, myocardial infarction, or sudden death), these findings have low predictive value in asymptomatic subjects.

Thallium myocardial perfusion scanning is a relatively new procedure for detecting CAD without cardiac catheterization. We examined whether combining ECG and thallium-scanning (TS) during maximal treadmill exercise would improve the predictive accuracy for detecting CAD *(Fleg et al., 1983b)*. Because cardiac catheterization cannot routinely be performed in asymptomatic individuals, especially in those with a normal response to these non-invasive tests for CAD, we monitored these subjects for the subsequent development of cardiac events (CE) indicative of CAD (angina pectoris, myocardial infarction, and sudden death).

Accordingly, 235 clinically healthy BLSA volunteers aged 40 to 92 years received ECGs and TS in conjunction with maximal treadmill exercise. The incidence of subsequent CE in each of four groups defined by the results of exercise ECGs (EE) and TS is shown in the table below.

Groups	*N*	*CE*	*Age*	*Follow-Up (Yr)*
−EE/−TS	179	3	59.1	2.5
−EE/+TS	22	0	58.3	2.8
+EE/−TS	20	1	65.4	2.7
+EE/+TS	14	5	69.3	3.2

The group with +EE/+TS was significantly older than the other three groups and had a markedly higher rate of CEs during the follow-up period. Both age and the combination of +EE/+TS were independent predictors of CEs. Coronary events consisted of angina pectoris in seven subjects and myocardial infarction and sudden death in one subject each.

In this asymptomatic population, the combination of an abnormal EE and an abnormal TS thus identified a group of elderly subjects with a strikingly high incidence of subsequent CE.

CHAPTER VII

BLSA: Acta and Agenda

INTRODUCTION

We have emphasized that this book is a progress report on the first 23 years of the Baltimore Longitudinal Study of Aging (BLSA), which has addressed a broad spectrum of questions about physiological, psychological, and psychosocial variables in the processes of aging. Despite its 23-year span, the bulk of its members have been followed for only about 13 years of their lives; since half of the participants were enrolled before they had reached age 50, substantial numbers of participants have not yet been followed into senescence and advanced old age (85 and older). The book is thus an interim report—a fact that receives further emphasis when we consider what we have learned and what still remains to be investigated.

INDIVIDUALITY AND SPECIFICITY OF AGING

Aging cannot be equated to a disease or disorder, nor can there be hope for a "magic bullet" that will cure or stop it. In the past 20 years most gerontologists have recognized the distinctly limited utility of a simple and sovereign notion of aging. Instead of hypothesizing a unitary or uni-causal process of aging, researchers have emphasized the need to consider the interacting influences of biological processes, personality and behavioral factors, social and environmental forces, and the idiosyncratic health behaviors and stresses of the individual. Similarly, the BLSA experience has made clear the complexity of aging processes. Instead of a single underlying mechanism, aging is now regarded as reflecting the expression of a host of processes that independently and in concert in the individual bring about the changes we recognize as aging. The BLSA experience has shown that multi-disciplinary longitudinal studies provide a fertile basis for generating and testing hypotheses regarding aging processes in men and women.

BLSA data indicate that aging is a highly individual process. Although cross-sectional observations show a significant decline in many physiological variables over the total age span, individual differences are very large. In some variables, individual 80-year-old subjects may perform as well as the average 50-year-old. Aging is highly specific not only for each individual but also for different organ systems within the same individual.

Because of the high degree of specificity of aging among different subjects and among different organ systems, chronological age itself is not a very reliable predictor of performance in individual adults. Recognition of this great diversity may prove of value in devising interventions to improve performance in normal aging subjects or for subgroups. Although no single "treatment" is likely to be discovered that will improve various kinds of performance in all or most people, a variety of different interventions or treatments tailored to critical personal characteristics might be developed that would be effective for those individuals.

The discrete character of various physiologic and behavioral functions has been confirmed by several analyses that have found no statistically significant factor common to them. The conclusions—that there is no single process of aging, and that so-called "physiological indices of aging" provide no better predictors of individual performance that chronological age—confirm the BLSA's choice of a multidisciplinary and multivariate approach, both over its 23-year history and in future studies that will address the problems it has identified.

PATTERNS OF CHANGE WITH AGE

From the mass of research results presented in Chapters V and VI, we can readily conclude that there is no one uniform age course for all variables. The evidence is conclusive that there are a variety of changes with age, of which we might identify six types or patterns. One pattern is stability, or the absence of any meaningful change with age in important functions or aspects of the person, ranging from resting heart rate to personality characteristics. The second pattern is characterized by declines with age that are due not to aging *per se* but to illnesses associated with age. Thus, although earlier studies reported a significant decline in plasma-testosterone levels with age, analyses performed on subjects who were carefully screened for diseases revealed no such age differences among healthy men. A third pattern shows steady declines in function in spite of good health or the absence of disease; creatinine clearance is a classic illustration of intrinsic change. Changes that occur precipitously in old age form a fourth pattern; such changes are often expressions of disease or closely related to disease, as in dementia. Other apparent changes may actually represent the body's attempts to maintain function with advancing age. This fifth pattern of change, exemplified in the posterior pituitary response to hyperosmolality and in the use of the Frank-Starling mechanism to maintain cardiac output during exercise, might be termed compensatory. The sixth type of change that occurs with the passage of time has little or nothing to do with age, or with health and disease, but reflects cultural changes that are of importance in interpreting research data on aging. A clear example of this is provided by the nutritional data indicating reduction of dietary cholesterol intake over the historical period during which BLSA data have been collected.

An intriguing opportunity for further research lies in those functions, predominantly psychosocial, that do not change with age, or change only late in life and in minor degree. The implications for both geriatric practice and social policy of the finding that personality is relatively unchanging, and that cognitive functions decline substantially only after age 70—and even then are in part stabilized by compensatory changes—must be studied both in the BLSA, in women as well as men, and in other populations. Perhaps the most important goal is an understanding of the underlying mechanisms responsible for both stability and change.

Complex physiological functions that require coordination among different organ systems show greater decrements with age than more simple responses. Some of the major impairments of aging may be due primarily to breakdowns in various regulatory mechanisms that reduce the adaptability of the individual, e.g., the age-related diminution in cardiovascular response to catecholamines. The effectiveness of any regulatory mechanism depends on its ability to detect the presence of significant deviations and to institute measures to counteract displacements. Both these mechanisms (loss of sensitivity and of responsivity) contribute to the phenomena of aging.

In its study of both physiological and psychosocial functions the BLSA has found evidence of mechanisms that in part compensate for declines that accompany aging. Examples range from cognitive to cardiac to aerobic function, and may result not only from voluntary actions by the individual (such as the cessation of smoking) but also from adjustments of which the individual is unaware (as when a deficit in memory is in part offset by increases in knowledge). Identification of the mechanisms underlying such compensatory adjustments might lead to the development of interventions that would enhance the quality of life of the elderly.

One of the fundamental endeavors of the BLSA has been to study the effects of aging as discrete from those of disease. Routine medical examinations have been supplemented by diagnostic procedures, such as stress testing, that make possible identification of diseases that would otherwise have remained occult. Screening for these illnesses allows investigators to study aging in large samples in the absence of clinically detectable illness. At the same time, longitudinal data from individuals who develop specific conditions can be used to follow the course of those diseases.

Analysis of BLSA longitudinal data indicates that a precipitous drop in any physiological or behavioral function is likely to be a manifestation of a pathological condition. A corollary is the hypothesis that, in variables that remain essentially stable over the adult life span, any significant change may be a manifestation of pathology. Further study of this hypothesis will be an important part of future BLSA analysis.

The work of the BLSA is just beginning. The establishment of a data base sufficiently large and exhibiting enough points of measurement over a period sufficiently long to make possible longitudinal, time-sequential, and cross-sequential analysis provides the ground for more sophisticated investigations of the questions it has already addressed, as well as of new questions that will arise in the process.

FUTURE DIRECTIONS

It is not possible to chart with any great degree of precision the specific directions of future research. Research directions are continuously under scrutiny and revision, and new investigators bring with them new methods and hypotheses. But we may sketch a few broad areas.

The failure of aging in some body systems to predict aging in other systems calls attention to the need both for more complete description of the course of aging and for more theoretical understanding of its mechanisms in each individual system. The intensive investigation of specific functions will thus remain an important part of the BLSA's mission.

The relative homogeneity of the BLSA population has made it possible for investigators to exclude from analysis factors or influences that show variation across demographic and socioeconomic parameters, and has speeded the identification of a variety of age-related patterns of stability and change. The validity of its findings for other populations, however, remains to be demonstrated. What we know now is to a great extent conditional on the subjects we have studied, the populations they are drawn from, the methods we have used, and the historical period in which the subjects have lived and been studied. Whether the patterns we have found will be repeated in other ethnic and racial groups, other historical periods, and other aspects of aging are studied remains unpredictable. In the near future one of the most significant aspects of generalizability will be tested, when longitudinal analysis of data on BLSA women will begin.

In the psychosocial area, another promising line of research will address the means of promoting health and independence in older individuals, particularly through research on the lifecourse patterns of health behaviors and their emotional, interpersonal, and experiential correlates. Future research is likely to emphasize the role of lifelong personality dispositions in shaping characteristic modes of perceiving health, maintaining fitness, and coping with illness. At the same time, research needs to consider the influence of life stresses and social supports on mental and physical health and functioning. Studies previously conducted on men will need to be extended to women, and sex differences (physiological, psychological, and social) must be examined for their possible significance in explaining differences between men and women in health and longevity.

As the BLSA engages new fields of research, and possibly new and more broadly representative populations—the longitudinal analysis of the data from the study of women is the most immediate example—it will be able to retest past findings and to determine the validity of new ones. Despite the conceptual and empirical gains won from our longitudinal and cross-sectional investigations, a great many questions remain to be answered. Only a few cohorts have been studied, and only over two decades of historical time. How other cohorts will age, and how other historical periods may influence the aging of its members, awaits further observations on this population as well as on others. In a society in which older people are exploring new roles and new solutions to old problems, the scientific and social importance of such study cannot be sufficiently emphasized. In short, the findings of the BLSA now constitute a kind of gerontological agenda for the future.

Appendix

Longitudinal Studies on the BLSA Population

Appendix

Longitudinal Studies on the BLSA Population

Arenberg, D., A Longitudinal Study of Problem Solving in Adults 5

Arenberg, D., Differences and Changes with Age in the Benton Visual Retention Test 14

Arenberg, D., Estimates of Age Changes on the Benton Visual Retention Test .. 21

Arenberg, D., Changes with Age in Problem Solving 25

Arenberg, D., Memory and Learning Do Decline Late in Life 40

Borkan, G.A., and Norris, A.H., Fat Redistribution and the Changing Body Dimensions of the Adult Male 51

Costa, P.T., Jr., Fleg, J.L., McCrae, R.R., and Lakatta, E.G., Neuroticism, Coronary Artery Disease, and Chest Pain Complaints: Cross-Sectional and Longitudinal Studies 71

Costa, P.T., Jr., and McCrae, R.R., Somatic Complaints in Males as a Function of Age and Neuroticism: A Longitudinal Analysis 79

Costa, P.T., Jr., and McCrae, R.R., An Approach to the Attribution of Aging, Period, and Cohort Effects 92

Costa, P.T., Jr., and McCrae, R.R., Concurrent Validation after 20 Years: The Implications of Personality Stability for Its Assessment 105

Costa, P.T., Jr., and McCrae, R.R., Personality as a Lifelong Determinant of Well-Being 129

Costa, P.T., Jr., McCrae, R.R., Andres, R., and Tobin, J.D., Hypertension, Somatic Complaints, and Personality 147

Costa, P.T., Jr., McCrae, R.R., and Arenberg, D., Enduring Dispositions in Adult Males 163

Costa, P.T., Jr., McCrae, R.R., and Arenberg, D., Recent Longitudinal Research on Personality and Aging 171

Costa, P.T., Jr., McCrae, R.R., and Norris, A.H., Personal Adjustment to Aging: Longitudinal Prediction from Neuroticism and Extraversion 215

Costa, P.T., Jr., and Shock, N.W., New Longitudinal Data on the Question Whether Hypertension Influences Intellectual Performance 223

Douglas, K., and Arenberg, D., Age Changes, Cohort Differences, and Cultural Change on the Guilford-Zimmerman Temperament Survey 234

Elahi, V.K., Elahi, D., Andres, R., Tobin, J.D., Butler, M.G., and Norris, A.H., A Longitudinal Study of Nutritional Intake in Men 245

Ensor, R.E., Fleg, J.L., Kim, Y.C., deLeon, E.F., and Goldman, S.M., Longitudinal Chest X-Ray Changes in Normal Men 264

page

Fleg, J.L., Das, D.N., and Lakatta, E.G., Right Bundle Branch Block: Long-Term Prognosis in Apparently Healthy Men 272

Hershcopf, R.J., Elahi, D., Andres, R., Baldwin, H.L., Raizes, G.S., Schocken, D.D., and Tobin, J.D., Longitudinal Changes in Serum Cholesterol in Man: An Epidemiologic Search for an Etiology 278

McCrae, R.R., and Costa, P.T., Jr., Aging, the Life Course, and Models of Personality 292

McCrae, R.R., Costa, P.T., Jr., and Arenberg, D., Constancy of Adult Personality Structure in Males: Longitudinal, Cross-Sectional, and Times-of-Measurement Analyses 304

Norris, A.H., Mittman, C., and Shock, N.W., Changes in Ventilation with Age . 311

Plato, C.C., and Norris, A.H., Osteoarthritis of the Hand: Longitudinal Studies . 318

Potter, J.F., Elahi, D., Tobin, J.D., and Andres, R., Effect of Aging on the Cardiothoracic Ratio of Men 325

Quilter, R.E., Giambra, L.M., and Benson, P.E., Longitudinal Age Changes in Vigilance over an Eighteen-Year Interval 331

Robertson-Tchabo, E.A., Arenberg, D., and Costa, P.T., Jr., Temperamental Predictors of Longitudinal Change in Performance on the Benton Revised Visual Retention Test among Seventy-Year-Old Men: An Exploratory Study 335

Rowe, J.W., Andres, R., Tobin, J.D., Norris, A.H., and Shock, N.W., The Effect of Age on Creatinine Clearance in Men: A Cross-Sectional and Longitudinal Study 345

Schlesselman, J.J., Planning a Longitudinal Study. I. Sample Size Determination 354

Schlesselman, J.J., Planning a Longitudinal Study. II. Frequency of Measurement and Study Duration 362

Shock, N.W., Energy Metabolism, Caloric Intake, and Physical Activity of the Aging 372

Shock, N.W., Andres, R., Norris, A.H., and Tobin, J.D., Patterns of Longitudinal Changes in Renal Function 384

Tobin, J.D., Physiological Indices of Aging 387

Tzankoff, S.P., and Norris, A.H., Longitudinal Changes in Basal Metabolism in Man 396

Journal of Gerontology
1974, Vol. 29, No. 6, 650-658

A Longitudinal Study of Problem Solving in Adults[1]

David Arenberg, PhD[2]

A longitudinal study of an initial sample of 300 men whose ages ranged from 24 to 87 years was carried out using modifications of the logical problems devised by John. Cross-sectional results for all problems showed that, with increasing age, proportions of subjects who solved decreased and uninformative inputs increased. The mean interval between first and second measures was 6.70 years. Longitudinal measures of change showed a mean decline in performance only for the group which was over 70 initially. Much of the decline was attributable to redundant inputs despite the fact that memory demands were minimized and the entire record of input-outcome events was always available. Six-year survival also was related to successful problem-solving performance in the first session.

ALTHOUGH reasoning is one of man's most cherished behaviors, and problem solving is a prevalent experimental approach to the study of reasoning in the laboratory, few cross-sectional studies of problem-solving performance and aging have been published; and to the writer's knowledge, no longitudinal study has been reported. In this paper, a longitudinal study of problem-solving performance is described. The initial sample consisted of 300 male volunteers in the Baltimore Longitudinal Study (see Stone & Norris, 1966) who ranged in age from 24 to 87 years at the time of their first problem-solving session.

In the brief review of the cross-sectional studies of problem solving and aging which follows, the studies in which the problems were similar to those used in this longitudinal study are discussed first. In those studies, problems were modifications of the type developed by John (1957). Jerome (1962) and Young (1966) used such problems and reported substantial age differences in performance between young and old adults. Similar problems were used in the present study and are described in greater detail in the Method section. The important point is that both studies showed that the old have considerable difficulty solving such problems.

Other cross-sectional studies have not consistently found age differences in problem-solving performance. In most of those studies, the problems required some form of concept identification. Wetherick (1964, 1966), in a number of studies of concept-type problem solving, has reported no age differences at least as often as age differences. Some of his negative results probably are attributable to his matching age samples on non-verbal intelligence which correlates with problem-solving performance. Such matching tends to select better performers in an old group than would be selected without matching. Even with the same samples, however, Wetherick (1964) has found age differences in some problems and not in others.

Age differences have not been found consistently in other concept-type studies. Wiersma and Klausmeier (1965) reported no age difference in time to solve four conjunctive problems with two relevant attributes. Although the number of incorrect identifications prior to solution increased monotonically with age, the authors reported no statistically significant age effect. It should be noted that the oldest group was between 35 and 51 years of age. Arenberg

---

[1] The author gratefully acknowledges the assistance of Don McQ Reynolds and Marcia H. Schwartz for administering and scoring problems, of Patricia Allen for administering them, and of Barbara Donovan and Karen Douglas for scoring and analyzing the data.

[2] Gerontology Research Center, NIH, Baltimore City Hospitals, Baltimore 21224.

(1968) found age differences in five variations of one-attribute concept problems. Brinley, Jovick, and McLaughlin (1974) also reported age differences in one-attribute problems in which both memory and organization were manipulated systematically. Young (1971) reported a study of concept identification in which she found consistent age differences for men in their 40s, 50s, and over 60; but the performance of their wives was not age related at any of the several levels of complexity used. The mean performance measure for the men over 60 was lower than the means for the younger groups for all six problem types. The women over 60 performed as well as the younger female groups on all problems.

In summary, the experimental literature on reasoning in gerontology indicates that: (a) few cross-sectional studies have been reported; (b) concept studies of aging have not always shown age differences; (c) logical problems like those devised by John (1957) have consistently shown age differences (Jerome, 1962; Young, 1966); and (d) no longitudinal study of problem solving has been reported.

The primary purposes of the present study are: (a) to determine whether performance on problems like those used by Jerome and by Young changes with age, and (b) to determine whether such changes are related to age, i.e., whether the old decline more than the younger subjects., The cross-sectional analyses also provide information about age differences in performance for rather large samples under conditions in which the subject has total control over the occurrence of each input-outcome event. In the previous studies the input-outcome events were paced, and the samples were small.

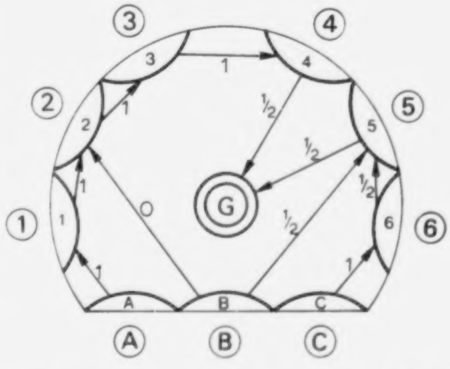
METHOD

Subjects.—All of the participants in the Baltimore Longitudinal Study who were available between 1962 and 1967 were included in this study. The resulting sample consisted of 300 male volunteers ranging in age from 24 to 87 at the time of the first administration of the problems. A vast majority of the men had some college education, and 81% of the sample had earned at least a bachelor's degree. The number in each decade with an earned degree is shown in Table 1. With very few exceptions, the participants were white, middle-class men employed (or retired) as scientists, professionals, or managers. WAIS Vocabulary scores ($\bar{X} = 66.8$) were above the national average and at least as high for the old as for the younger groups. For all analyses, the subjects initially in their 20s were combined with the group in their 30s, and those in their 80s were combined with the group in their 70s.

Procedure.—In all problems a display of ten lights was used (see Fig. 1), and in each problem a set of logical relations between lights was defined by the experimenter. The lights which were related were connected by arrows on a problem disk displayed with the lights. Each arrow was assigned one of several possible meanings. Three possible meanings were used in this study: (a) effector, (b) combinor, and (c) preventor. The effector meaning was a simple cause-and-effect relation between two lights. The arrows on the sample problem disk which were effectors are designated "1" in Figure 1. (Of course, the specific designations did

Table 1. Mean WAIS Vocabulary Scores and Number with One or More Earned Degrees.

	20s	30s	40s	50s	60s	70s	80s
Vocabulary	60.8	63.1	67.3	66.4	67.9	69.0	69.4
No. with earned degree	4	29	60	66	43	37	4
N	5	40	79	80	47	45	4



O = Preventor ½ = Combinor 1 = Effector

Fig. 1. Sample problem. The solution sequence is:

$$\begin{array}{l} A \longrightarrow 1 \longrightarrow 2 \longrightarrow 3 \longrightarrow 4 \\ \qquad\qquad\quad C \longrightarrow \left.\begin{array}{l}6\\ \\ B\end{array}\right\} \longrightarrow 5 \end{array} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \longrightarrow G$$

where A, B, and C represent lights which were activated by button presses, and 1, 2, 3, 6, 4, 5, and G represent lights which resulted as outcomes of the previous input.

not appear on the disk displayed to the subject.) For example, the arrow from 1 to 2 meant that whenever the input was 1, 2 resulted. In the published studies in which similar problems had been used, input-outcome events occurred at a fixed pace. Lights which were on became inputs and were replaced by their outcomes at the end of a constant period of time (typically 3 sec.). In the present study, no external pace was imposed. The transition from an input to an outcome was totally under the control of the subject. When he activated 1, for example, 2 would result only after the subject indicated that he wanted the results of his input. Then the experimenter would extinguish 1 and activate 2 using switches on the control unit. With 2 lighted, the subject had three options: (a) he could clear the board (with a large button in the upper right corner of his display panel not shown in Fig. 1) and resume; (b) he could request the results of 2, i.e., 2 becomes an input; or (c) to the 2 already lighted he could add other lights, and the next input would consist of 2 and the additional lights. The other effector arrows operated in a similar way. In the sample problem (Fig. 1), A resulted in 1, 2 resulted in 3, 3 resulted in 4, and C resulted in 6.

The combinor meaning is designated "1/2" on pairs of arrows with common targets in Figure 1. For example, the arrow from 4 to G and the arrow from 5 to G were a pair of combinor arrows. Operationally, an input of 4 without 5 (or 5 without 4) resulted in a blank outcome. The target light, G, resulted only if the input consisted of both 4 and 5. Similarly, 5 resulted from an input of both B and 6; neither B without 6 nor 6 without B resulted in 5.

The preventor meaning is designated "0" in Figure 1. The arrow from B to 2 on the sample disk was a preventor arrow. Operationally, 2 did not result whenever an input included B. In other words, B prevented 2; B alone resulted in a blank outcome, 1 alone resulted in 2, and B with 1 in the same input resulted in a blank outcome.

The three possible meanings of arrows were explained to the subject in great detail using the sample problem. Each light on the periphery of the disk could be activated by pressing a corresponding button adjacent to that light. The goal light, G, in the center of the disk had no corresponding button on the subject's panel. In the sample problem and in all other problems, G could be lighted only as an outcome. During the sample problem subjects were shown that each of the peripheral lights could be lighted directly by pressing its button, but the goal light could only be lighted as a result of a particular input. In the sample problem, only if an input included both 4 and 5 would G result.

The task in each problem was to arrive at the outcome, G, via a sequence of inputs without pressing the buttons for the numbered lights. It was strongly impressed upon every subject that the buttons for numbered lights could be pressed to obtain information about the meanings of specific arrows (and that such information was important in solving a problem logically); but only the buttons for lettered lights could be used in a solution sequence of inputs.

After all the meanings of arrows in the sample problem had been explained, the solution sequence was demonstrated. This was done after the subject understood that both 4 and 5 were necessary in the final input, and that it was necessary to achieve 4 and 5 together as an input without pressing the buttons for 4 or 5 (or any other numbered light). He was shown how to obtain the outcome, 4; how to obtain the outcome, 5; and how to obtain the outcome, 4 and 5, which could then be used as the final input to reach the outcome, G. As indicated in Figure 1, the solution sequence was: (a) press A which yields outcome, 1; (b) add nothing and 1 yields outcome, 2; (c) press C which added to 2 yields outcome, 3 and 6; (d) press B which added to 3 and 6 yields outcome, 4 and 5 (3 yields 4, and at the same time B together with 6 yields 5); and (e) add nothing and 4 together with 5 yields outcome, G.

The sample problem was used to introduce the many rules and procedures for solving this type of problem. They included demonstrating input-outcome transitions, emphasizing that at the subject's command (and not until then) outcome lights replaced input lights, explaining the three possible meanings of arrows, showing that inputs consist of lights which can result either from button pressing or from an outcome of the previous input, explaining and demonstrating the restrictions on button pressing during solution sequences (but not during collection of information about specific arrow meanings), and illustrating a system for recording all input-outcome events to provide a written record which could be reviewed at any time.

Subjects were instructed to attempt to minimize the number of inputs required to solve the problems but were informed that it was more important to reach a solution using many inputs than not to reach a solution at all. They were also informed that time was a secondary consideration in evaluating performance and that no savings in time could compensate for even one superfluous input.

Following the sample problem, the practice problem was administered. The experimenter instructed the subject to make specific inputs and explained the information available after each outcome. During this part the subject pressed buttons and recorded inputs and outcomes. After all arrow meanings could be identified, he was given the opportunity to synthesize independently a solution sequence of inputs. Questions about rules and procedures were encouraged, and the subject was told at the beginning of this and all subsequent problems that he could ask such questions at any time. If the subject needed help solving the practice problem, the experimenter attempted to elicit the source of the difficulty in order to maximize the subject's understanding of the task and permissible operations.

The necessity of maintaining an accurate record of all inputs and outcomes was stressed. It was explained that such a record relieved the subject of the burden of remembering his previous inputs and their outcomes, so that his major efforts could be applied to reasoning.

All problems were administered during a single half-day session for each subject. Between 45 and 120 min. were required for instructions, the sample problem, and the practice problem. Up to 150 min. were used to work on one to three problems in the main set. Typically, two breaks of about 10 min. were provided; the first usually followed the practice problem. All subjects attempted the first problem and also attempted the second and third problems if time permitted. Each problem in the main set was logically more complex than the preceding problem. The three problems were logically identical to the first three levels used by Jerome (1962).

Because the apparatus was not automated, the experimenter was in the room with the subject at all times. As a fortuitous result, questions about rules and procedures which arose during problems (even in the main set) could be answered and misunderstandings could be clarified.

The second session was scheduled during a subsequent visit to the Gerontology Research Center a minimum of 6 years after the first session. The mean interval was 6.70 years, and mean intervals for all decades were quite similar (ranging from 6.57 to 6.83 years). Of the 300 men who participated initially, 224 returned. The number of men who did not return and the major reasons are shown in Table 4 for each decade (age at first session). Problems in the main set of the second session were logically identical to the comparable problems in the first session. In all other respects, the second session was the same as the first.

Dependent measure.—The primary purpose of the study was to measure age changes in effectiveness in reasoning, i.e., in analyzing and synthesizing information in solving logical problems. The level of difficulty of the problems was selected so that few of the participants in the Baltimore Longitudinal Study would be unable to solve at least the first problem without making it trivial. In order to assess on-going problem-solving performance, a measure was developed to reflect the potential gain in information of each input. This measure was the number of uninformative inputs. Each input prior to achieving a solution sequence could be evaluated as (potentially) informative or uninformative depending upon the pool of information available at that point in the subject's solution. The number of uninformative inputs was the primary dependent measure; and these inputs were further classified as overtly redundant, directly inferable, or indirectly inferable in order to elucidate the age deficits in reasoning that were hypothesized.

RESULTS

Problem I: Cross-sectional.—Of the total of 300 men in the initial sample, 263 solved the first problem. The proportion of men who successfully solved the problem is shown for each age group in the top part of Table 2. The proportions declined monotonically with age; a substantially higher proportion of men under

Table 2. Proportions Who Solved Problems I, II, and III.

	Under 40	40s	50s	60s	Over 70	All
Problem I						
First session (N)	45	79	80	47	49	300
—proportion solved	.96	.94	.91	.79	.73	.88
Returned (N)	34	65[a]	65	34	26	224[a]
—proportion[b] solved 1st session	.94	.97	.94	.79	.81	.91
—proportion[b] solved 2nd session	.97	.98	.95	.85	.82	.94
Problem II						
First session (N)	45	77	76	40	34	272
—proportion solved	.87	.92	.87	.65	.71	.83
Returned (N)	34	65	62	30	20	211
—proportion[b] solved 1st session	.88	.92	.90	.60	.70	.84
—proportion[b] solved 2nd session	.94	.95	.87	.71	.69	.87
Problem III						
First session (N)	39	66	56	21	20	202
—proportion solved	.95	.91	.91	.86	.65	.89
Returned (N)	29	58	46	15	12	160
—proportion[b] solved 1st session	.97	.90	.91	1.00	.67	.91
—proportion[b] solved 2nd session	.94	.95	.91	.65	.80	.89

[a] Three additional subjects returned, but data could not be obtained. [b] Of those who returned, proportion who solved.

Table 3. Mean Number of Uninformative Inputs.

	Under 40	40s	50s	60s	Over 70	All
Problem I						
Solved 1st session						
—number	43	74	73	37	36	263
—mean	5.12	7.35	6.88	8.46	8.50	7.17
Solved both sessions						
—number	32	63	59	24	16	193
—mean 1st session	5.28	6.71	7.07	6.96	6.62	6.61
—mean 2nd session	4.00	3.90	5.39	5.08	11.37	5.14
—mean change	−1.28	−2.81	−1.68	−1.87	4.75	−1.47
Problem II						
Solved 1st session						
—number	39	71	66	27	24	227
—mean	7.36	10.35	10.65	11.00	14.00	10.40
Solved both sessions						
—number	30	59	50	17	10	166
—mean 1st session	8.07	8.93	9.86	11.24	12.90	9.53
—mean 2nd session	4.67	7.92	6.88	8.18	19.90	7.77
—mean change	−3.40	−1.02	−2.98	−3.06	7.00	−1.77
Problem III						
Solved 1st session						
—number	37	60	51	18	13	179
—mean	8.89	9.12	10.31	16.28	14.08	10.49
Solved both sessions						
—number	27	52	40	11	6	136
—mean 1st session	9.74	8.02	8.20	20.36	15.33	9.74
—mean 2nd session	10.11	4.56	9.82	13.55	7.83	8.08
—mean change	0.37	−3.46	1.62	−6.82	−7.50	−1.65

60 than over 60 successfully solved the problem.

The mean number of uninformative inputs for successful solutions in each age decade are shown in Table 3. (Note that the higher the number of uninformative inputs, the less effective the solution.) There was a general increasing trend with the largest increments between the 30s and 40s and between the 50s and 60s. The product-moment correlation between uninformative inputs and age was 0.12 ($p < 0.05$, $N = 263$). Of the three kinds of uninformative inputs, both overtly redundant inputs and directly inferable inputs contributed substantially to the high mean of total uninformative inputs for the men over 60.

Before considering longitudinal analyses of age changes, it is necessary to determine the effects of attrition (due to failure to return) on initial cross-sectional data. Of the 300 men in the original sample, 224 returned for another administration of logically equivalent problems. The returning subsample of 224 men had a higher proportion of solvers (91.7%) in the first session than the subsample which did not return (77.6%). Another effect of the attrition was that age differences in the mean number of uninformative

Table 4. Number and Reasons for Failure to Return.

	Under 40	40s	50s	60s	Over 70	All
Total	11	11	15	13	23	73
—died*	2,0	1,0	4,1	2,1	6,6	15,8
—withdrew*	9,0	8,1	8,2	8,1	6,2	39,6
—ill*	0,0	1,0	0,0	0,1	3,0	4,1

* Entries separated by comma indicate number who solved and number who did not solve Problem I.

inputs based on all subjects who solved Problem I were not found for the groups who returned and successfully solved this problem in both sessions; e.g., subjects in their 40s had a mean of 6.71 uninformaitve inputs, whereas the men over 70 had a mean of 6.62 (see Table 3).

Problem I: Longitudinal.—For those subjects who solved the first problem in both sessions, cross-sectional age differences in uninformative inputs were apparent the second time. Inspection of the mean change for each decade (see Table 3) indicates that all groups improved except the men initially over 70. The decline in reasoning effectiveness (increase in uninformative inputs) was larger for the 70-year group than for each of the other age groups (all $ps < 0.05$, Mann-Whitney test). More than half of the men in each group below 70 improved; but only 5 of the 16 men over 70 improved.

Among the 49 men over 70 in the original sample, 12 died before they were due to return (see Table 4), and survival was related to performance. Of the 36 men who solved the first problem initially, 83% survived, whereas only 54% of the 13 men who failed that problem survived ($p < 0.05$, one-tailed exact probability). This relationship was not found for the vocabulary measure. A comparison of mean WAIS Vocabulary raw scores of those who died and those who lived showed no difference ($\bar{X}_{\text{Died}} = 69.3$; $\bar{X}_{\text{Lived}} = 68.9$). It should be pointed out that the vocabulary measure typically was obtained 1 to 3 years prior to the problem-solving session. The relationship between survival and problem-solving success was not attributable to age; i.e., survivors were not predominantly the younger men in the over-70 group who were also more frequently successful in solving Problem I. For the 36 solvers in this group, the mean age of survivors was 74.3 years, and of nonsurvivors, 74.2 years; for the 13 nonsolvers, the mean age of survivors was 75.4 years, and of nonsurvivors, 76.9 years. As would be expected, deaths among the men who were between 50 and 70 initially were far less frequent, but even in this age range the results were similar. Of the men in their 50s and 60s who successfully solved Problem I, 5% died; whereas 12% of those who failed to solve that problem died.

A second and third problem were administered in each session if time permitted. The attrition in the sample due to insufficient time was a source of bias in the performance measures for these problems. Prior performance of subjects who did not reach a problem was typically poorer than for those who reached that problem.

Problem II: Cross-sectional.—Of the 300 men in the original sample, 272 reached Problem II and had valid performances. The age distribution of those who solved that problem is shown in Table 2. As with Problem I, the proportions of successful solutions for Problem II declined with age.

The measure of reasoning effectiveness for those who solved Problem II also was related to age. The correlation between uninformative inputs and age was 0.17 ($p < 0.01$, $N = 227$). Table 3 shows that the largest increments in uninformative inputs occurred between the 30s and 40s and between the 60s and 70s.

Of the 224 men who returned for a second session 6 (or more) years after their first session, 13 had not reached Problem II the first time; and of the 211 who had attempted to solve that problem, 178 (84%) had been successful. The distribution by age groups can be seen in Table 2. Of the 209 men who reached that problem the second time, 182 (87%) solved successfully; this included three men who solved Problem II the second time but had not reached that problem the first time.

Problem II: Longitudinal.—There were 166 men who solved Problem II both times and who could be included in an analysis of change in number of uninformative inputs. Unlike Problem I, the initial performance of the subsample who solved Problem II both times was age-related (see Table 3). A similar result was obtained for the second measure of uninformative inputs. Just as in Problem I, all age groups except the oldest showed small mean improvements; but the oldest group showed a substantial age decline in reasoning effectiveness. In Problem II, that decline was almost entirely attributable to an increase in the number of overtly redundant inputs.

Problem III: Cross-sectional.—Of the 300 men in the original sample, 202 reached Problem III and had valid performances, and 179 (89%) solved it successfully. The distribution by age groups of those who solved is shown in Table 2. Similar to Problems I and II, the proportions of successful solutions for Problem III declined with age. For successful solutions, the number of uninformative inputs was also age related ($r = 0.20$, $p < 0.01$, $N = 179$). The largest age difference in reasoning effectiveness occurred between the 50s and 60s (see Table 3).

Sixty-four of the 224 men who returned and could be scheduled for a second session had not reached Problem III in the first session. Of the 160 men with valid performances and who returned, 145 (91%) had successfully solved that problem the first time; the age distribution can be seen in Table 2. Of the 179 men who reached Problem III the second time, 160 (89%) solved it successfully.

Problem III: Longitudinal.—A total of 136 men solved Problem III both times and could be included in a longitudinal analysis of change in number of uninformative inputs. The groups over 60 had substantially more uninformative inputs than the groups under 60 the first time. This finding is similar to the results of the cross-sectional analysis of the entire original sample. As can be seen in Table 3, the relationship between reasoning effectiveness and age was not monotonic the second time; and as a result, the age changes were not consistent with the previous data. Unlike the results of Problems I and II, the group initially in their 70s did not decline in performance; but it should be noted that due to many sources of attrition, that sample was reduced to 6, and was an atypical subsample of their age peers.

"Nongoal-directed" behavior.—Early identification of the subgoal condition (the necessary input for the goal outcome, G) was an essential component of the backward solution strategy which was emphasized in Jerome's (1962) study. Jerome found, however, that his old subjects typically did not identify the subgoal condition early. A measure was devised to quantify such "nongoal-directed" behavior in the present study. This measure is the number of inputs which were prior to identification of the subgoal condition, but which were not directly related to the goal light. In Table 5, it can be seen that this measure increased across the age dimension for all three problems (first session).

Table 5. Mean Number of Early Inputs Not Directly Related to Goal.

	Under 40	40s	50s	60s	Over 70	All
Problems						
I (N = 263)	1.81	1.82	1.81	1.81	2.92	1.97
II (N = 227)	1.02	1.68	2.61	1.96	3.08	2.02
III (N = 179)	1.43	1.25	2.16	2.39	1.62	1.69

Discussion

The cross-sectional results based on the initial sample showed an unequivocal decline in reasoning performance with age. Although the pattern of age differences varied somewhat, all three problems showed a decrease in proportion of successful solutions and a decline in effectiveness of solutions with age. In the first problem, the decline in reasoning effectiveness was partially attributable to an increase in overtly redundant inputs; in the second and third problems, the declines were almost totally attributable to increases in overtly redundant inputs. Despite the substantial differences in procedures, these results are consistent with the two-sample age differences reported by Jerome (1962) and by Young (1966).

A measure of "nongoal-directed" behavior (the number of inputs which were prior to identification of the subgoal condition, but which were not directly related to the goal light) showed an increasing monotonic relationship with age. These cross-sectional results are consistent with Jerome's (1962) findings that the old did not identify the subgoal condition early as part of the backward solution strategy which was emphasized in his instructions. In the present study no explicit strategy was included in the instructions; nevertheless, "nongoal-directed" inputs early in the solution increased with age. Perhaps the old are poor at formulating and adopting an effective strategy as well as poor at carrying out such a strategy when it is explicitly taught to them.

In the longitudinal analyses, mean age changes were limited to the subjects who were initially in their 70s. For these men the number of uninformative inputs increased, and the increases were primarily attributable to more overtly redundant inputs. The memory component of the task had been minimized by having subjects maintain written records of every input-outcome event. Despite the availability of such a record the oldest subjects repeated inputs frequently, and it was an increase in these repeated inputs that accounted for a major portion of the decline in the effectiveness of their reasoning performances. Descriptively this took the form of additional attempts at solution sequences which were variations of earlier attempts, but which were ineffective because they did not make use of the information available from previous input-outcome events.

Rarely are studies of age changes free of methodological problems, and this study has its share. Ideally, a time-sequential design would have been used with a minimum of two cohorts at each age. Sometimes early and late entries into a sample can be separated to create different cohorts for time-sequential analyses, but in this study the sample was too small and the time span of data collection was too short. Also it would have been preferable to have data for all three problems for all subjects in both sessions, but time was a restricting factor which biased the proportions-solved measure, and both time and ability to solve a problem were biasing selective factors for the reasoning-effectiveness measure. Furthermore, subjects who returned were not representative of the initial sample. All of these factors affected the data positively, i.e., by excluding a subsample of predominantly poorer performers. Nevertheless, age changes in reasoning performance were obtained for the group initially in their 70s, and that is the primary finding of this study. The positive biases mentioned above could have masked age changes for men below 70, but the age changes for the 70-year-olds emerged despite these biases.

The results of a follow-up analysis of mortality and reasoning performance for the 49 men in the initial group over 70 showed a relationship between survival and success in solving Problem I. Similar results of so-called "terminal drop" have been reported for several psychological performance measures (see Granick, 1971, for a review of the relationship of cognitive performance and mortality). The mortality results of particular interest in this study were the differential relations between death and vocabulary on the one hand and death and reasoning on the other hand. Death occurred before the second administration of the problems for 17% of subjects over 70 who had successfully solved Problem I, but for 46% of those men who had failed that problem. On vocabulary performance, however, the mean of survivors was the same as the mean of those who died.

All the results together suggest strongly that the age changes found in reasoning effectiveness among the men over 70 who successfully solved problems in both sessions were not early manifestations of "terminal drop." These men were healthy enough to return, they were able to solve problems in the second session, and their vocabulary performance did not decline. If the decrements in performance are early indices of "terminal drop," however, then decline in reasoning performance should predict death; i.e., men whose performance declined from the first to the second problem-solving session should have a less favorable survival experience than those whose performance was maintained or improved. Time (and follow-up data) will tell.

SUMMARY

A longitudinal study of logical problem solving was conducted. The initial sample consisted of 300 men, predominantly well-educated and middle-class, who ranged in age between 24 and 87 years. Six years later, 224 men returned and attempted to solve logically equivalent problems. Cross-sectional analyses of age differences and longitudinal analyses of age changes in proportions who solved successfully and in effectiveness of successful solutions resulted in the following:

1. Age differences favoring the young were found in proportions of successful solutions in all three problems, and the largest differences occurred between groups under 60 and groups over 60.

2. For the successful solvers, age differences favoring the young were found in effectiveness in attaining solutions in all three problems.

3. Proportions of successful solutions in the second session (6 or more years after the first) were quite similar to comparable proportions in the first session.

4. For the men who successfully solved a specific problem in both sessions, age declines in reasoning effectiveness were found only for the group over 70 (initially) in Problems I and II.

5. Of the 49 men in the over-70 group in the first session, 12 died before they were due to return 6 years later. Survival was related

to success in solving Problem I but was not related to WAIS Vocabulary. Of the 36 men who solved the problem, 17% died; of the 13 men who failed to solve the problem, 46% died. The vocabulary means for survivors ($\bar{X} = 68.9$) and for those who died ($\bar{X} = 69.3$) were virtually identical.

Consistent age differences were found across the entire adult age range, but longitudinal age changes were not found for the age groups below 70. Several positive biases were operating which probably reduced the true age changes: those who returned were a select subsample of the original sample; those who reached Problems II and III were, for the most part, superior performers; and only those who successfully solved a particular problem in both sessions could be included in that analysis of change in reasoning effectiveness. Despite these biases, age changes were found for the group over 70, evidence that ability to solve reasoning problems declines in late life even for a highly select group of men.

REFERENCES

Arenberg, D. Concept problem solving in young and old adults. *Journal of Gerontology*, 1968, *23*, 279-282.

Brinley, J. F., Jovick, T. J., & McLaughlin, L. M. Age, reasoning, and memory in adults. *Journal of Gerontology*, 1974, *29*, 182-189.

Granick, S. Cognitive aspects of longevity. In E. Palmore & F. C. Jeffers (Eds.), *Prediction of life span.* Lexington, MA: D. C. Heath & Co., 1971.

Jerome, E. A. Decay of heuristic processes in the aged. In C. Tibbitts & W. Donahue (Eds.), *Social and psychological aspects of aging.* New York: Columbia Univ. Press, 1962.

John, E. R. Contributions to the study of the problem-solving process. *Psychological Monographs*, 1957, *71*, 1-39.

Stone, J. L., & Norris, A. H. Activities and attitudes of participants in the Baltimore Longitudinal Study. *Journal of Gerontology*, 1966, *21*, 575-580.

Wetherick, N. E. A comparison of the problem-solving ability of young, middle-aged and old subjects. *Gerontologia*, 1964, *9*, 164-178.

Wetherick, N. E. The inferential basis of concept attainment. *British Journal of Psychology*, 1966, *57*, 61-69.

Wiersma, W., & Klausmeier, H. J. The effect of age upon speed of concept attainment. *Journal of Gerontology*, 1965, *20*, 398-400.

Young, M. L. Problem-solving performance in two age groups. *Journal of Gerontology*, 1966, *21*, 505-509.

Young, M. L. Age and sex differences in problem solving. *Journal of Gerontology*, 1971, *26*, 330-336.

Journal of Gerontology
1978, Vol. 33, No. 4, 534-540

Differences and Changes with Age in the Benton Visual Retention Test[1]

David Arenberg, PhD[2]

The relation of adult age and performance on memory for designs (Benton Revised Visual Retention Test) was determined for men by analyzing: (a) three cross-sectional samples (Ns = 402, 162, and 293); (b) two longitudinal samples (repeated measures at least 6 years apart, Ns = 268 and 82); and (c) within-cohort comparisons of men born in the same period, but tested at different times. The results were essentially the same for the age differences, age changes, and estimates of age changes based upon within-cohort differences. Increases in errors were small for the young groups, moderate for the men in their 50s and 60s, and substantial for the men over 70. WAIS Vocabulary measures for these same samples showed small cross-sectional differences favoring the older men, no overall longitudinal change but small relative deficits for the older participants, and small declines in estimates of age changes based upon comparisons of independent samples born during the same period. In general, the results indicate age declines in memory-for-designs performance for men particularly late in life, but only small (if any) age declines in vocabulary for the same samples. No correlation was found in either longitudinal sample between change in memory performance and change in vocabulary score.

ALTHOUGH memory loss is considered a pervasive problem of the aged, few cross-sectional studies of memory and aging have been reported outside the verbal-memory domain, and not one longitudinal study of age changes in nonverbal memory has appeared in the gerontological literature. In this study, performance consists of reproducing geometric designs from memory. In addition to cross-sectional data, age changes based on conventional longitudinal data and estimates of age changes based on comparisons of independent samples from the same birth cohort measured at different times are included.

Age differences in memory for designs have been reported across the adult age range (Benton, 1963; Davies, 1967; Graham & Kendall, 1960; Poitrenaud & Clément, 1965) and between young and old adults (Arenberg, 1977). The results of those studies are consistent; mean errors increased with increasing age.

The data in the current study are from the Baltimore Longitudinal Study (see Stone & Norris, 1966) during the period from late 1960 to mid-1973. It was predicted that: (a) errors would increase with age cross-sectionally; (b) age changes would be found longitudinally; and, (c) these longitudinal changes would increase with age. The fact that the data were collected over an extensive period of time provided an opportunity to explore whether estimates of age changes based upon age differences within birth cohorts (performance for men in an early sample compared with performance of men born at the same time but from a later sample) would be consistent with the longitudinal age changes.

METHOD

Subjects

The subjects included every participant in the Baltimore Longitudinal Study for whom a measurable initial performance during the period 1960-1973 was available (N = 857). The participants in the Baltimore Longitudinal Study are males between the ages of 18 and 102, but heavily concentrated between the ages of 30 and 80. They are all volunteers who agree to come to Baltimore City Hospitals periodically for 2½ days during which an extensive variety of physiological, biochemical, and behavioral measures are obtained.

---

[1] Many psychologists participated in the collection and scoring of the data in this study. The contributions of Patricia Lamb, Robert Walker, Richard Mathias, Don Reynolds, Barbara Donovan, Patricia Allen, Marcia Schwartz, Steven Kanis, Darrell Gray, and Marian Hedrick are gratefully acknowledged. I am particularly indebted to Karen Douglas not only for administering and scoring, but especially for programming and analyzing the data.

[2] Gerontology Research Center, National Institute on Aging, NIH, Baltimore City Hospitals, Baltimore 21224.

The men are predominantly white, well-educated, and of high socioeconomic status living in the Baltimore-Washington area.

Procedures

For purposes of analysis, the total sample of 857 was divided into three subsamples on the basis of the date of the initial measure. The periods were: (a) early, 1960.9 - 1964.9, (b) middle, 1965.0 - 1968.5, and (c) late, 1968.6 - 1973.5.

The primary cross-sectional analysis was based on the early sample with the middle and late samples serving as replications. The primary longitudinal analysis was based on the "repeats" from the early sample. A small longitudinal replication was based on the "repeats" from the middle sample. In addition, estimates of age changes were based upon within-cohort comparisons (i.e., age differences between subgroups born at the same time but measured at different times). These birth cohort subsamples were from the early and late samples.

All participants were given Form C of the Benton Revised Visual Retention Test, Administration A (Benton, 1963), usually on their first 2½-day visit. Form E was administered on a subsequent visit 6 (or more) years after the first test. A form is made up of ten designs. Each of the first two designs consists of one major (geometric) figure; each of the other eight designs consists of two major figures and a peripheral, minor figure. Under Administration A, the standard procedure, each design is displayed for 10 sec and withdrawn. The subject's task is to reproduce the design from memory at his own pace. Errors were scored according to the test manual (Benton, 1963). Each design was scored by two psychologists independently, and the infrequent disagreements were resolved by discussion or a third psychologist. The dependent variable was the number of errors in all ten reproductions.

Whenever either form of the Benton was administered, the Vocabulary subtest of the Wechsler Adult Intelligence Scale was also administered. Every analysis of the memory data was paralleled by the same analysis of the vocabulary data. In addition, the correlations between change in memory performance and change in vocabulary were calculated for the two longitudinal samples.

RESULTS

Cross-Sectional

The three cross-sectional samples were classified into seven age groups, and the means for Benton errors are shown in Table 1. The monotonic increases in means were consistent with product-moment correlations between errors and age of .47 (N = 402), .47 (N = 162), and .51 (N = 293) in the early, middle, and late samples respectively (all *ps* < .01).

Longitudinal

Longitudinal analyses were based upon all participants in the early and middle samples for whom a valid second measure was obtained 6 (or more) years after the first measure. The mean interval between measures was 6.7 yr for the early sample and 6.5 yr for the middle sample. These samples were grouped according to age at first measure. The means for first and second measures and their differences are shown in Table 2.

Table 1. Mean Errors for Seven Age Groups in Three Cross-Sectional Samples[a]

	Early		Middle		Late	
Age groups	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$
≥ 80	8	11.75	3	12.00	12	8.33
70s	55	6.33	20	6.05	50	6.64
60s	66	4.58	35	5.09	39	5.31
50s	100	3.50	37	4.51	48	4.48
40s	98	2.88	40	3.15	41	3.76
30s	67	2.61	15	2.73	61	3.13
< 30	8	1.25	12	2.75	42	2.55

[a] More detailed descriptive data are available from the author upon request.

Table 2. Mean Errors for First and Second Measures for Two Longitudinal Samples.

	Early				Middle			
Initial age	N	1st $\bar{X}$	2nd $\bar{X}$	Diff. $\bar{X}$	N	1st $\bar{X}$	2nd $\bar{X}$	Diff. $\bar{X}$
≥ 80	1	13.00	11.00		1	11.00	10.00	
70s	24	6.33	9.33	3.00	8	5.50	8.75	3.25
60s	45	4.93	5.53	.60	17	4.94	5.59	.65
50s	77	3.36	3.97	.61	22	4.32	4.64	.32
40s	70	2.70	2.99	.29	26	3.08	3.00	-.08
30s	48	2.65	3.00	.35	3	4.67	3.00	
< 30	3	.67	67		5	2.80	2.00	

The mean increase in errors in the early sample was .68; together with a correlation of .60 (ignoring age) between the first and second measures and a sample size of 268, that mean change was statistically significant ($p < .01$). In the middle sample, the mean increase was .39 errors; together with the correlation of .41 (ignoring age) between the first and second measure and the sample size of 82, that change was not statistically significant. However, the magnitude of change increased with age in both samples. The correlations between age at first measure and increase in errors was .20 ($N = 268$, $p < .01$) for the early sample and .27 ($N = 82$, $p < .01$) for the middle sample.

In order to determine the relationship of age with that part of the second measure which was independent of the first measure, part correlations (age with residual of second measure on first; see DuBois, 1957) were calculated; these correlations were .36 ($p < .01$) for the early sample and .45 ($p < .01$) for the middle sample.

The relationship of magnitude of change with age is depicted in Fig. 1. This figure shows cumulative distributions of age changes for five age groups in the early sample. Most striking is the marked difference between the oldest group and the other four. For example, over 50% of the men in each of the four youngest groups did not decline, but only 28% of the group initially over 70 did not decline. An increase of more than three errors was found in fewer than 10% of the two youngest groups and in fewer than 20% of the groups initially in their 50s and 60s, but almost 50% of the men over 70 declined to that extent.

Typically in longitudinal studies of cognitive performance, the participants who return for repeated measures are a positively biased subsample of the initial sample (e.g., Riegel et al., 1967). The effects of attrition in the two longitudinal samples can be seen by comparing the mean for each age group in the initial sample (in Table 1) with the mean of the first measures for that age group (in Table 2) for the subsample for whom a second measure was obtained. In the early sample, 268 of the original 402 subjects had a valid second measure. The "repeat" age groups performed with about the same or fewer errors than the initial age groups. In the middle sample, of the 162 subjects in the initial sample, 82 had a valid second measure and returned in time to be included; except for the groups below 40, the "repeat" age groups committed fewer errors than the initial age groups. No statistically significant difference was found, however, between "repeats" and "nonrepeats" in the two samples combined ($F < 1$, unweighted means ANOVA for unequal cell frequencies). Reasons for attrition are shown in Table 3 by age decade for the early and the middle samples.

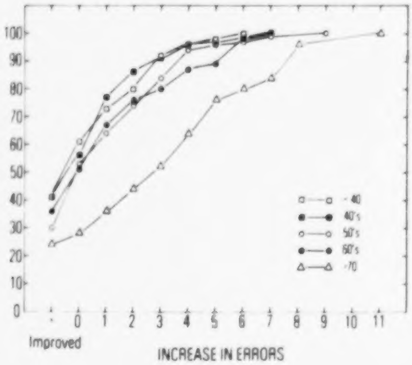


Fig. 1. Cumulative percentages of magnitude of change in five age groups (early sample).

Table 3. Reasons for Nonrepeats.

	< 30	30s	40s	50s	60s	70s	80s	Total	% of Initial Sample
Early sample									
Withdrew[a]	4	12	24	17	15	11	0	83	21
Died	0	2	2	4	3	14	6	31	8
Other[b]	1	5	2	2	3	6	1	20	5
Total	5	19	28	23	21	31	7	134	33
Middle sample									
Withdrew[a]	3	4	3	7	6	4	1	28	17
Died	0	0	1	3	4	7	1	16	10
Other[b]	4	8	10	5	8	1	0	36	22
Total	7	12	14	15	18	12	2	80	49

[a] "Withdrew" includes failure to respond (to request to schedule a visit).

[b] "Other" includes visual impairment, refusal, invalid measure, scheduling omission during a visit.

Table 4. Mean Errors and Age for Nine Birth Cohorts at Two Periods of Measurement — Independent Samples.

	Early[a]			Late[a]		
Cohorts	N	Age	Errors	N	Age	Errors
1877 - 1884	11	81.1	7.91	2	87.8	9.00
1885 - 1892	53	72.8	5.70	10	82.4	8.20
1893 - 1900	52	65.9	5.25	49	73.2	6.59
1901 - 1908	55	57.6	3.29	35	65.5	5.49
1909 - 1916	99	50.2	3.49	33	56.6	4.97
1917 - 1924	77	41.8	2.71	40	49.2	3.67
1925 - 1932	46	34.6	2.43	31	41.9	3.81
1933 - 1940	6	27.8	1.50	52	33.6	3.21
1941 - 1948	1	20.8	2.00	40	27.3	2.32

[a] Two subjects in the early sample were born before 1877, and one subject in the late sample was born after 1948.

Birth Cohort

Although the Baltimore Longitudinal Study was not designed to test differences within cohorts, it was possible to compare the early sample with the late sample for groups of men born during the same period. In Table 4, the same nine 8-year birth periods are listed for each sample, and the mean errors and mean ages are tabulated. For each of the nine birth cohorts, the error mean was larger for the late sample (the older subsample) than for the early sample (the younger subsample). Although comparisons of means for groups of similar age consistently show more errors for the late sample (e.g., 7.91 and 8.20 errors for ages 81.1 and 82.4 in Table 4), these time-lag differences account for only a small part of the within-cohort estimates of age declines.

Means for the seven cohorts born between 1877 and 1932 in the early sample are shown in Fig. 2 connected by a dotted line. On the same figure, mean changes are shown for participants (from six of those seven cohorts) who had two valid measures. These changes are depicted by solid triangles connected by dashed lines. Similar to the shape of the cross-sectional curve, age changes were small for the youngest groups, moderate for the groups with mean ages initially of 57 and 65, and substantial for the oldest cohort in their late 60s or early 70s at the time of their first test. Estimates of age changes are shown for these same six birth cohorts by comparing men born during the same periods, but tested at two different times. The cross-sectional points (open circles) from the early sample are connected by solid lines to the cross-sectional points (filled circles) from the late sample for men born during the same periods. All of these within-cohort comparisons indicated age differences with the largest estimate of age changes in the earliest born (oldest) cohort.

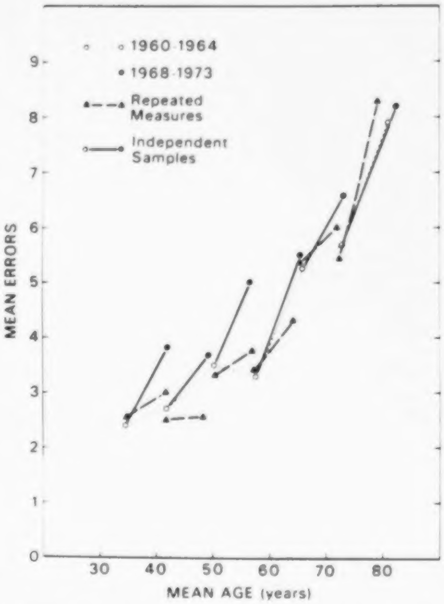


Fig. 2. Cross-sectional, longitudinal, and within-cohort comparisons for groups classified by period of birth and time of measurement.

WAIS Vocabulary

The WAIS Vocabulary test was included in the Baltimore Longitudinal Study primarily as a marker variable to describe the high level of the study group. A vocabulary measure was available for virtually every Benton performance included in this report. As a result, it is possible to analyze the Vocabulary measures for cross-sectional age differences, longitudinal age changes (repeated measures), and estimates of age changes based upon independent cohort subsamples born during the same period, but tested at different times.

In all three cross-sectional samples, the relationship between age and vocabulary raw score was slightly positive; $r = .14$, .07, and .13 in the early, middle, and late samples, respectively. Although the correlations were small and statistically significant only for the early ($p < .01$, N = 400) and late ($p < .05$, N = 293) samples, the consistency of the positive trend indicates that the older men were slightly higher in vocabulary performance than the younger men.

Longitudinally, a mean age change was not found for either the early or the middle sample. For the early sample (N = 267), the means for the first and second measures were 66.96 and 66.84 ($t < 1$); for the middle sample (N = 82), the means were 66.01 and 65.52 ($t < 1$) (the correlations between measures were .86 and .87 despite the restricted range). However, magnitude of change was related to age; the means for the older groups tended to decline slightly, whereas the means for the younger groups increased slightly. The correlations between age and change were −.19 for the early sample ($p < .01$) and −.06 for the middle sample ($p > .05$). Part correlations between age and that part of the second measure which was independent of the first measure (age vs residual) were −.16 ($p < .01$) for the early sample and −.13 ($p > .05$) for the middle sample.

Within-cohort comparisons of WAIS Vocabulary measures showed a small overall mean difference favoring the early sample ($\bar{X} = 66.5$) over the late sample ($\bar{X} = 64.5$) (unweighted means ANOVA, $F(1,566) = 4.98$, $p < .05$). However, the estimates of age changes were not age related; that is, the within-cohort differences were similar for all age groups ($F < 1$).

The correlations between change in memory performance (increase in errors) and change in vocabulary (increase) were .01 (N = 267) for the early sample and −.09 (N = 82) for the middle sample. Neither correlation was statistically significant ($t < 1$). The correlations between initial vocabulary score and change in Benton performance (increase in errors) were −.09 for the early sample and −.04 for the middle sample. Neither correlation was statistically significant ($t < 1$).

DISCUSSION

Age differences in the three cross-sectional samples were quite similar, and these results are consistent with the increases in errors reported in previous cross-sectional studies of memory-for-designs. The longitudinal age changes were remarkably similar to the cross-sectional age differences. Errors increased little at the younger ages, moderately for the groups tested initially in their 50s and 60s, and substantially for the men in their 70s. This result was found for two independent longitudinal samples, one initially measured between 1960 and 1964, the other between 1965 and 1968.

The longitudinal findings were supported by the estimates of age changes based upon comparisons of independent samples from the same birth cohort measured at different times. Again the largest differences (between subsamples born at the same time) were found for the earliest born (oldest) cohorts.

Different results were found for the Vocabulary subtest of the WAIS with virtually the same subjects. Cross-sectional comparisons showed small age differences favoring the older men. Mean longitudinal changes were virtually nil, but magnitude of change was related to age. The oldest men declined slightly, and the youngest men improved slightly. However, change in vocabulary was unrelated to change in Benton performance. The men who declined in memory performance were not the men who declined in vocabulary.

The Benton task appears to fall under the M dimension (ability to reproduce immediately information presented either auditorily or visually) in Horn's (1975) description and discussion of intelligence and aging, and the WAIS Vocabulary task clearly is in the Gc dimension (crystallized intelligence). The memory results of this study are consistent with previous cross-sectional findings of memory for geometric figures (Benton, 1963; Davies, 1967; Graham & Kendall, 1960; Poitrenaud & Clément, 1965) and of other memory tasks which Horn characterizes as M measures (see Botwinick & Storandt, 1974). The longitudinal age changes found in the present study (supported by the estimates of age change based upon within-cohort comparisons) add the important information that Benton performance declines within the individual, especially late in life. These findings are strong evidence for maturational change in memory for geometric designs and suggest that previous cross-sectional age differences in performance on other tasks in the M dimension of intelligence are also attributable to maturational declines.

In the later of the two longitudinal samples, the initial memory performance of those men who returned was slightly superior to that of those who did not return. However, in the first and larger longitudinal sample, returnees and nonreturnees were not different in their initial performance. Even if there were biases due to attrition (superiority of those who return), they should be positive biases; that is, they should operate to reduce age changes.

All of the results point to the same conclusion. Memory-for-designs performance declines with age, particularly after age 70.

Although memory-for-designs performance is intended to measure nonverbal memory, it is clear that at least some geometric figures can be verbalized, and such verbalizations can be used to encode designs during presentation and can serve as retrieval cues at the time of reproduction. It has been demonstrated in this laboratory that elderly men with less education than the participants in the present study benefited substantially from auditory description of the Benton designs during visual presentation (Arenberg, 1977). It is likely that the older men in the present study did not use verbal encoding maximally. Thus, some part of their decline in performance may well be remediable by training. It would be valuable to determine whether such training at the time of initial performance would prevent or substantially reduce declines especially in those over 70. If the effectiveness of the training is not maintained, it would be important to determine whether the decrement was attributable to failure to apply the encoding techniques or to reduced effectiveness of the encoding presumably due to age changes in the brain.

Even among the group over 70 when initially tested, some men did not decline. Why do some men decline while others do not? This is an important research question, and one approach is to look for correlates of change. When other information is available, it may be possible to characterize those who do not decline. If such characterizations can be validated on other groups, directions for preventive intervention to maintain performance may be indicated. Currently the medical and physiological data available for the participants in the Baltimore Longitudinal Study are being explored in search of correlates of changes in performance on memory for designs.

SUMMARY

From 1960 to 1973, as part of the Baltimore Longitudinal Study, the Benton Revised Visual Retention Test (Benton, 1963) was administered initially to 857 men between 18 and 102 yrs of age. When this group was divided into an early, middle, and late sample, the correlations between number of errors and age were .47 (N = 402), .47 (N = 162), and .51 (N = 293), respectively (all *p*s < .01). Repeated measures for the early and middle samples were obtained 6 (or more) years after the initial measures. The correlations between change in errors and age were .20 (N = 268) and .27 (N = 82) (both *p*s < .01); and the part correlations between age and residual of second mesure on first were .36 and .45 (both *p*s < .01). The subsamples of men who returned performed slightly better initially than their age peers who did not return, but the differences were not statistically significant.

Men in the early sample were compared with men in the late sample who were born during the same period. The within-cohort comparisons are estimates of age changes, and they were consistent with the longitudinal measures of age changes. The largest within-cohort difference was found for the earliest born (oldest) birth-cohort.

Results were quite different for WAIS Vocabulary scores from the same samples. Cross-sectionally the correlations between vocabulary and age were positive; that is, the older men had slightly higher scores initially than the young men. Overall changes measured longitudinally were virtually nil, but magnitude of change was correlated with age. The old men declined slightly and the young men improved slightly. Within-cohort comparisons of vocabulary scores did not show a relation between age and estimate of age change. The correlations between change in Benton performance and change in vocabulary score were not statistically significant.

Clearly memory for designs declines with age in men, substantially for men initially over 70. The age changes in vocabulary were

small. The lack of relationship between change in memory for designs and change in vocabulary is further evidence against a general decline in cognitive performance with age.

REFERENCES

Arenberg, D. The effects of auditory augmentation on visual retention for young and old adults. *Journal of Gerontology*, 1977, *32*, 192-195.

Benton, A. L. *The Revised Visual Retention Test: Clinical and experimental applications*. 3rd ed. Psychological Corp., New York, 1963.

Botwinick, J., & Storandt, M. *Memory, related functions and age*. Charles C Thomas, Springfield, IL, 1974.

Davies, A. D. M. Age and memory-for-designs test. *British Journal of Social & Clinical Psychology*, 1967, 6, 228-233.

DuBois, P. H. *Multivariate correlational analysis*. Harper, New York, 1957.

Graham, F. K., & Kendall, B. S. Memory-for-Designs Test: Revised general manual. *Perceptual & Motor Skills*, 1960, *11*, 147-188.

Horn, J. L. Psychometric studies of aging and intelligence. In S. Gershon & A. Raskin (Eds.), *Geriatric psychopharmacology: The scene today*. Raven Press, New York, 1975.

Poitrenaud, J., & Clément, F. La détérioration physiologique dans le Test de Rétention Visuelle de Benton: Résultat obtenus par 500 sujets normaux. *Psychologie Française*, 1965, *10*, 359-368.

Riegel, K. F., Riegel, R. M., & Meyer, G. A study of the dropout rates in longitudinal research on aging and the prediction of death. *Journal of Personality & Social Psychology*, 1967, *5*, 342-348.

Stone, J. L., & Norris, A. H. Activities and attitudes of participants in the Baltimore Longitudinal Study. *Journal of Gerontology*, 1966, *21*, 575-580.

Journal of Gerontology
1982, Vol. 37, No. 1, 87-90

Estimates of Age Changes on the Benton Visual Retention Test[1]

David Arenberg, PhD[2]

Performances on the Benton Visual Retention Test were collected for 894 men 17 to 96 years of age between 1960 and 1976 and were analyzed using slopes to estimate time-of-measurement effects. For each of the nine birth cohorts and eight age groups, performance (errors) was plotted on calendar time. Comparisons of the within-birth-cohort slopes showed an increase in slope with age (rho = 0.92); earlier born (older) cohorts had larger slopes indicating increased errors per year. Comparisons of within-age-group slopes showed no difference among age groups. Slopes of four of the five oldest cohorts were larger ($p < .05$) than the common age-group slope. Whether the common age-group slope was a secular effect, a cohort effect, or some combination of the two, the data were consistent with age changes in performance late in life confirming longitudinal results of direct measures of change previously reported.

Key Words: Aging, Human memory, Cognition, Estimates of age change, Time-of-measurement effects

RECENTLY, Arenberg (1978) reported substantial mean age changes on the Benton Visual Retention Test (Benton, 1963) over a 6-year interval only for men over 70 when initially tested. Mean differences within birth cohorts were included also and were consistent with the direct measures of change; the largest mean difference between men tested early (younger) and men tested later (older) in the study was found for the earliest born cohort.

Although the study was not designed for sequential analyses (see Schaie, 1977), there was some suggestion in the data that at least part of the within-cohort differences (and perhaps part of the longitudinal changes) were due to time-of-measurement effects within age groups. The purpose of this paper is to examine more intensively the time-of-measurement effects within birth cohorts and within age groups: (1) to obtain estimates of age changes within many cohorts and to determine whether these estimates were related to age; (2) to obtain estimates of nonmaturational changes within many age groups and to determine whether these estimates were equivalent for all age groups; and (3) to determine whether the within-age-group effects could account for the within-birth-cohort effects. Regression analyses were used rather than mean differences because the data were collected continually rather than within narrow time periods separated by a long time interval. In these regression analyses, time is a continuous rather than a discrete variable.

METHOD

Participants. — The data included in these analyses were collected from 1960 to 1976 from participants in the Baltimore Longitudinal Study of Aging. They were unpaid volunteers who agreed to come to Baltimore City Hospitals periodically for 2½ days during which an extensive variety of physiological, biochemical, and behavioral measures were obtained. The men were predominantly white, educated, and of high socioeconomic status living in the Baltimore-Washington area. The sequential analyses were based upon 894 men with a first-time measure on the Benton Revised Visual Retention Test during the first 16 years of the study (between late 1960 and late 1976). These 894 men include the 692 from the cohort comparisons of means in the previous report, 161 men tested between the beginning of 1965 and mid-1968, and 41 men tested between mid-1973 and the end of September, 1976. The men ranged in age from 17 to 96 years at the time they were tested.

---

[1] Many experimenters were involved in administering and scoring the Benton VRT since 1960, most recently Barbara Hiscock, Susan Goldstein, and Judy Friz. Their contributions are gratefully acknowledged. I am especially indebted to Judith Plotz for programming as well as for administering and scoring the test.

[2] Gerontology Research Ctr., National Institute on Aging, NIH, Baltimore City Hospitals, Baltimore, MD 21224

Procedures. — The procedures were described in detail in the previous report (Arenberg, 1978). Form C of the Benton Revised Visual Retention Test was administered to virtually every participant in the Baltimore Longitudinal Study. Administration A was used; this is the standard procedure in which each design is displayed for 10 sec, and the participant can take as much time as he needs to reproduce the design from memory.

The dependent measure was the number of errors for all 10 designs. The designs were scored independently by two psychologists according to the test manual, and the infrequent disagreements were resolved by discussion or by a third psychologist.

Analysis. — Regression analyses were carried out for nine birth cohorts and for eight age groups. Each regression analysis was based on all the men born within a specific 8-year period (birth cohort) or within an 8-year age span (age group). First a scatter plot was made of all the points within each birth cohort and within each age group. A point represents one man's first Benton performance (number of errors) and calendar time of its occurrence. For a specific birth cohort, the slope of errors on calendar time is an estimate of linear age change for that cohort. In addition, slopes for age groups were calculated to determine whether there were linear nonmaturational effects over the 16-year period.

RESULTS

An example of a birth-cohort plot of errors and calendar time is shown in Fig. 1. These points represent all men born between the beginning of 1885 and the end of 1892 whose first Benton measure was obtained between 1960 and 1976. The men in this cohort were 68 to 75 years old at the beginning of the study (late 1960) and 84 to 91 at the end of the 16-year period (late 1976). The slope of the best-fit (least squares) straight line for this birth cohort was 0.35 (i.e., there was an average increase of 0.35 errors per year).

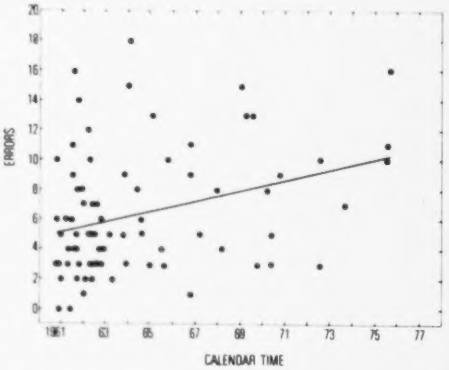


Fig. 1. Scatter plot of Benton errors and calendar time for 78 men born between 1885 and 1892.

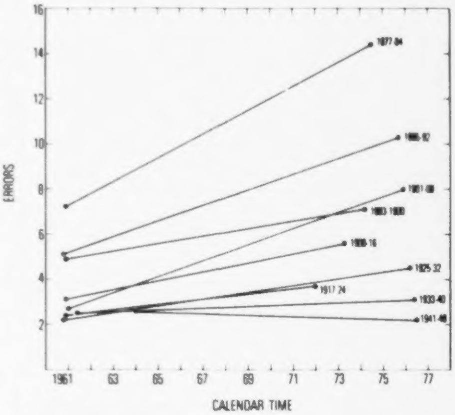


Fig. 2. Regression lines of Benton errors and calendar time for nine birth cohorts.

The regression lines for all nine birth cohorts are shown in Fig. 2. The rank-order correlation (Spearman rho) between birth cohort and slope was 0.92 ($p < .01$); the earlier born, the larger the estimate of age change. A test of common slope rejected the null hypothesis, F (8,871) = 2.53, $p < .05$; the slopes were significantly different.

When the men were categorized by age group rather than by birth cohort, the regression lines of errors on calendar time were calculated and are shown in Fig. 3. The test for common slope did not reject the null hypothesis, F (7,866) = 1.07, $p > .05$; the common slope was 0.074 which was different from zero, F (1,866) = 10.13, $p < .01$). The age groups did not differ in slope, but the common slope was different from zero.

If we assume that the time-of-measurement effect within age groups is a secular effect, that is, some systematic effect during the 16 years of the study, then part of the estimates

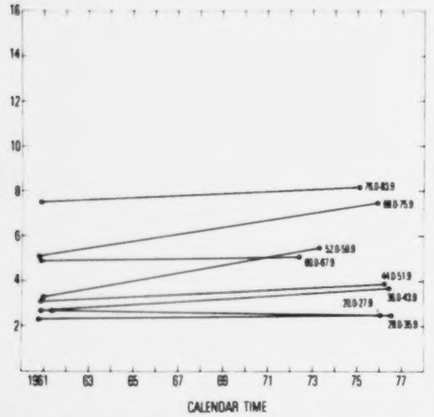


Fig. 3. Regression lines of Benton errors and calendar time for eight age groups.

of age changes based on the analyses of birth cohorts is attributable to secular rather than maturational effects. The common age-group slope was compared with the slope for each birth cohort to determine whether that estimate of age change was significantly different from the estimate of secular change. (In order to make these statistical comparisons of independent data, the points in the birth cohort were omitted from the calculation of the common age-group slope for each comparison.) Four birth-cohort slopes were significantly larger ($p < .05$) than the common age-group slope: 1877-1884, 0.49; 1885-1892, 0.35; 1901-1908, 0.35; and 1909-1916, 0.20.

DISCUSSION

These data are quite consistent with the longitudinal (repeated-measures) findings reported previously. Not only were mean errors related to age, but the estimates of age change were related to age. The rank-order correlation between mean age of cohort and estimate of age change was extremely high (0.92). Furthermore, for four of the five oldest birth cohorts, the estimates of age change were significantly larger than the estimate of secular effects.

The rationale for comparing the common age-group slope with each within-cohort slope was based on the assumption that the age-group slope is an estimate of the nonmaturational effects over time and each birth-cohort slope represents a combination of those effects and the maturational effect for that cohort. The common age-group slope, however, could be due to a systematic cohort effect (i.e., within each age group, the men born later committed more errors than the men born earlier). If we assume that the common age-group slope is due to cohort differences rather than a secular effect, then two conclusions can be drawn: (1) the cross-sectional age differences underestimate the contributions of maturation because the cohort differences were in the opposite direction, namely, later born commit more errors than earlier born; and (2) all of the within-cohort effects are maturational because there were no secular effects. Therefore, the more the age-group slopes were due to cohort differences, the more the cross-sectional and within-cohort results must represent maturational decline. Whether the time-of-measurement effect within age-groups was purely secular, purely cohort, or some combination of the two, the conclusions are the same; the data are totally consistent with maturational declines late in life with the magnitude of decline related to age.

One other possibility that could account for the time-of-measurement effect within age groups is sampling. If the caliber of men in each age group decreased over the period of the study, then errors would increase with time. This, however, could not account for the increases in errors for the oldest birth cohorts being significantly larger than the increases within age groups. Sampling should affect age groups and birth cohorts to the same degree. Therefore, the maturational explanation fits here as well.

The regression method developed to quantify linear effects over time within cohorts and within age groups identified time-of-measurement effects in both types of analyses. It was possible to compare not only magnitudes of time-of-measurement declines among birth cohorts and among age groups, but with these data, the method provided comparisons of individual birth cohorts with a time-of-measurement effect common to all age groups. In that way, it was possible to conclude that the time-of-measurement declines for the older cohorts, which were substantially larger than the common age group declines, are attributable to maturation.

In toto, the results are quite consistent with the longitudinal findings. The evidence is clear; on average, the older the man, the poorer his memory-for-designs performance at any one time and the steeper the rate of decline.

REFERENCES

Arenberg, D. Differences and changes with age in the Benton Visual Retention Test. *Journal of Gerontology*, 1978, *33*, 534-540.

Benton, A. L. *The Revised Visual Retention Test: Clinical and experimental applications* (3rd ed.). Psychological Corp., New York, 1963.

Schaie, K. W. Quasi-experimental research designs in the psychology of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. Van Nostrand Reinhold, New York, 1977.

Craik, F.I.M.; and Trehub, S. (Eds.),
Aging and Cognitive Processes.
New York: Plenum, 1982

Changes with Age in Problem Solving

David Arenberg

Gerontology Research Center
National Institute on Aging
Baltimore, Maryland

Reasoning is among the most cherished of man's abilities. It is, however, an aspect of cognitive performance which has proved more difficult to study than several others such as intelligence, memory, and learning. As a result, the reasoning literature is more meager and less systematic than the literature in other areas; and, not surprisingly, that picture is reflected in the area of reasoning and aging as well.

Typically, problem solving is used to study reasoning quantitatively. Sometimes performance is measured by the number of problems solved correctly; at other times the problems are designed so that finer aspects of the solutions can be quantified. An extensive review of problem-solving performance and aging can be found in Rabbitt (1977), and subsequent research was reviewed by Giambra and Arenberg (1980). The most recent review of the area is by Denney (1982).

As in other areas of cognitive performance and aging, most problem-solving research is cross-sectional. Many of these studies show age differences with old groups performing less well than young adults. It is recognized among cognitive investigators in aging that in a cross-sectional study age and birth cohorts are confounded and, therefore, age differences may reflect cohort differences and/or age changes. Group differences found in cross-sectional studies are sometimes referred to as "age/cohort" differences.

Longitudinal studies provide direct measures of change within individuals, but such studies are not without their problems. In a longitudinal study, age is confounded with time-of-measurement effects; that is, something may be happening during the course of the study which could account for or contribute to nonmaturational performance changes within individuals over time. Attrition is also a problem in longitudinal studies. Some participants leave the study for various reasons including death and declining health. It is unlikely that those who drop out are representative of the initial study group. As a result, longitudinal samples may be biased; typically, the better performers continue in the study.

Sequential strategies have been designed to avoid or minimize some of these problems (see Schaie, 1977), but they are not panaceas. These strategies require independent samples of several birth cohorts and several age groups at different times of measurement. Sequential analyses can provide group estimates of age changes to confirm longitudinal findings and can also provide estimates of nonmaturational changes over time.

The data segment of this chapter consists of 13 years of an ongoing longitudinal study of concept problem solving carried out in the Baltimore Longitudinal Study of Aging (BLSA). Several types of analyses of number of problems solved correctly are included: (1) cross-sectional age differences; (2) longitudinal changes following a six-year interval; (3) estimates of age change for many birth cohorts; and (4) estimates of nonmaturational change over time for many age groups. In addition, some longitudinal analyses are reported for measures of effectiveness in reaching solutions. This segment is followed by some comments on the current status and future directions of problem solving and aging.

Data collection in the longitudinal study of concept problem solving was initiated in 1967. At that time, two general types of problem-solving tasks had been used in more than one cross-sectional study of aging. Jerome (1962) and his colleague, Young (1966), had shown clear and consistent age differences in performance on complex logical problems developed by John (1957). Cross-sectional results from a longitudinal study using modifications of such problems (see Arenberg, 1974) were quite consistent with the age differences reported by Jerome and by Young. In those problems, the task is to light a goal light by a sequence of button presses involving other lights which are logically related. Arrows indicate which lights are logically related. Subjects are instructed about the possible relations that arrows represent. Each arrow indicates the direction of the relation, but the specific relation must be determined by the subject. In the studies by Jerome and Young, solutions required appropriately sequenced inputs within rather narrow time periods. The major modification introduced in Arenberg's longitudinal study was to make the task subject-paced.

The other general type of task used in studies of aging was concept problem solving. Wetherick's studies (1964, 1966) were of that kind. Unlike the studies with problems of the type used by John, consistent age differences in performance had not been found in the concept studies. It should be pointed out that in some of Wetherick's studies the old and young groups had been matched on nonverbal intelligence, and this almost certainly reduced age differences that may have existed. In 1967, therefore, when this longitudinal study was initiated, even the cross-sectional picture of problem solving and aging was not clear, and no data were available at all to answer the longitudinal question, whether reasoning changed with age. At that time, gerontologists were becoming increasingly more aware that cross-sectional studies are not dependable portrayals of change in an individual (or a group) with age even in the area of intelligence where many such studies had been reported. One purpose of this longitudinal study was to provide descriptive, quantified measures of change in problem-solving performance among an elite group of men.

As will be seen in the results thus far, cross-sectional age differences are quite apparent. The longitudinal results are not nearly so clear cut, but they are probably attenuated by biases due to attrition. The estimates of age change based on sequential analyses are showing substantial aging effects only for the oldest birth cohort. No evidence of nonmaturational changes over time is emerging.

The Longitudinal Study of Concept Problem Solving

In addition to gross measures involving number of problems solved correctly, the problems were designed to provide measures of how effectively each problem was solved. One way to investigate effectiveness is to allow subjects to select the instances. Then each selection can be evaluated for the amount of information gained. There was also the possibility that the pattern of selections would help us understand how a subject goes about solving problems. Typically, in concept problems, the experimenter selects the sequence of instances; this is referred to as a reception paradigm. In that way, the amount of information available to a subject can be controlled. Obviously this has great appeal; experimenters like to maintain tight control. It ensures that all subjects have the same amount of information at each point in the problem. When subjects are permitted to select instances, the general procedure referred to as the selection paradigm, that kind of control is relinquished. Furthermore, under typical conditions, logically equivalent selections can result in substantially different amounts of information gain. A procedure described by Arenberg (1970) was developed to minimize that problem in a selection paradigm.

In all of the problems in this study, there are four binary dimensions. The four dimensions are labeled A, B, C, and D; and the values for each dimension are 1 and 2. Therefore, the two attributes of dimension A are A1 and A2, the two attributes of B are B1 and B2 and so forth. Each instance consists of one attribute from each dimension, for example, A1, B2, C2, D1. The subject's task is to identify the concept by selecting instances which the experimenter designates as positive or negative. Using this information, the subject attempts to identify the concept with as few selections as he can. That is the solution to the problem.

The problems are presented as poisoned-food problems. Previous experience with less educated subjects had shown that some people have much difficulty understanding even simple concept problems when abstract dimensions are used. By using the language of poisoned foods, these difficulties were substantially reduced. In the language of poisoned foods, the attributes are foods, the selections (or instances) are meals, the concepts (solutions) are poisoned foods, and the designations are "Died" (positive) and "Lived" (negative).

So, in simple (one-attribute) problems, subjects are told that one of the eight foods has been poisoned and their task is to discover that poisoned food. In order to obtain information to solve the problem, they select meals. Each meal consists of one A food, one B food, one C food, and one D food. Each meal is designated "Lived" or "Died" by the experimenter. Whenever the poisoned food is included in a meal, the experimenter says, "Died"; meals which do not include the poisoned food are designated "Lived." Subjects are instructed to solve the problem with as few meals as they can but not to offer a solution until they are certain they have enough information to solve the problem.

Six different types of problems are used in this study, and a subject attempts to solve two of each type, 12 problems in all. Two types are one-attribute problems, that is, the concept is defined by one of the eight attributes (A1, A2, B1, B2, C1, C2, D1, D2). The only difference between these two types of simple, one-attribute problems is that in one the designations are predominantly positive and in the other they are predominantly negative.

The other four types are two-attribute problems; two are conjunctive and two are disjunctive. In the conjunctive problems, subjects are told that there are two attributes which together define the concept. Any instance which includes both of the attributes is a positive instance, an exemplar of the concept. Any instance which does not include both attributes (one or both are missing) is a negative instance, a nonexemplar of the concept. Subjects are told that two foods are poisoned but that a meal is fatal only if both are included. One type of conjunctive problem has low initial information, that is, the first instance is designated negative. Regardless of the selection, the designation is "Lived." This eliminates 6 of the 24 possible solutions. The other type of conjunctive problem has high initial information, that is, the first instance is designated positive ("Died"). This eliminates 18 of the 24 possible solutions.

Similarly, there are two types of disjunctive problems—one with high initial information and one with low initial information. In a disjunctive problem, a concept has two defining attributes, but an instance is designated positive if either attribute (or both) is included. An instance is designated negative only when neither attribute is included. Subjects are told that two foods are poisoned and that either is fatal. In a disjunctive problem with high initial information, the first instance is designated negative ("Lived"); this eliminates 18 of the 24 possible solutions. In a disjunctive problem with low initial information, the first instance is designated positive ("Died"); this eliminates 6 of the 24 possible solutions.

With binary dimensions, conjunctive and disjunctive problems are virtually mirror images of each other. A positive designation in one is equivalent to a negative designation in the other. Logically they are identical (and that helps when we use computers to administer and analyze these problems). As a result, it is possible to compare performance on conjunctive and disjunctive problems of the same type (i.e., high with high initial information and low with low). Conceptually, however, "bothness" and "neitherness" sometimes seem different and frequently are dealt with differently by problem solvers.

In order to minimize the memory component of the task, subjects are required to write every selection and its designation throughout each problem. They are also encouraged to write notes whenever they believe that would be helpful. Furthermore, the entire procedure is subject-paced. Subjects decide when to make a selection and when to identify a solution. They are given as much time as they need to solve each problem. While they are solving a problem, the concept rule for that problem type is prominently displayed. Any procedural question is answered by the experimenter. The six problems in a block are presented in four different orders, but the two one-attribute problems always precede the four two-attribute problems. For each subject, the six problems in the second block are presented (after a rest of about 10 minutes) in the same order as the first block.

Participants in the Baltimore Longitudinal Study of Aging are predominantly white, educated men[1] living in the Baltimore–Washington area and employed (or retired) as managers, scientists, or professionals (see Stone & Norris, 1966). The study group is sometimes described as "self-recruited," because of the fact that most participants were recruited by men already in the study. The BLSA was initiated in 1958 with a small group recruited by Dr. W. W. Peter. The men spend two and one-half days (every year or two) at the Gerontology Research Center in Baltimore participating in many physiological, biochemical, and behavioral studies.

Results

The data were collected from 1967 through 1979. First are the analyses of the gross measure of number of problems solved correctly. These begin with cross-sectional data, proceed in some detail with the two-point (at least six years apart) longitudinal data, and end with some regression analyses based on first-time data to provide estimates of age changes and estimates of nonmaturational changes over time. (These are analogs of Schaie's cross-sequential and time-sequential analyses but use slopes instead of mean differences because the data are collected continually.) Then some longitudinal analyses of the effectiveness measures are presented.

---

[1] Beginning in 1978, women were included in the Baltimore Longitudinal Study; however, only the data for the men are reported here.

Cross-sectional Data. Table 1 shows the cross-sectional means of number of problems correct for the seven age groups from the 20s to the 80s; the numbers of subjects in each age group (N) are also given. The table shows that there is a monotonic decrease from youngest to oldest.

Longitudinal Data. By the end of 1979, 376 men had returned for a repeat session. Of these, 87% (327) had a correctness measure for all twelve problems the second time. These are the men whose longitudinal correctness data are presented below.

Table 2 shows the mean number of problems correct for the first and second sessions for seven age groups. For this select group of men who continued in the Baltimore program and were able to attempt all 12 problems both times, the means decrease monotonically both times. The mean change measures, however, are not so consistent; nevertheless, the mean changes are positive for the youngest groups and negative only for the men who were in their 60s or 70s when first measured. The relationship between age and magnitude of change can be seen more directly using the correlation of age and individual change ($r_\Delta = -.08$). It is interesting to note that this relationship between age and magnitude of change occurs only in the second block. For the six problems in the first block, the age correlation with change in number of problems correct is only $-.028$; for the same six problems in the second block, the age correlation is $-.094$ ($p < .05$). It is possible that switching back to the types of problems previously encountered in the first block is more interfering for the older men after six years of aging.

Smaller groups of logically identical problems can be examined in the same way to determine whether particular types of problems contribute to these changes in overall performance with age. In the four simple (one-attribute) problems, the decreases with age are small, but they are virtually monotonic at both times. For this select group of subjects, however, magnitude of change is marginally related to age ($r_\Delta = -.07$).

Table 1. Cross-Sectional Means—Number of Problems Correct

	N	Mean
20s	71	10.4
30s	143	9.7
40s	154	9.0
50s	142	8.2
60s	118	8.0
70s	101	6.6
80s	22	5.4

Table 2. Longitudinal Means—Number of Problems Correct

	N	Mean 1st	Mean 2nd	Change
20s	10	10.3	10.8	0.5
30s	31	10.1	10.2	0.1
40s	96	8.9	9.2	0.4
50s	86	8.4	8.5	0.1
60s	69	8.5	8.1	−0.4
70s	31	6.9	6.7	−0.2
80s	4	3.5	4.5	—
All	327	8.6	8.6	0.1

The complex problems with high initial information tend to be easier than their low-initial-information counterparts. The typical monotonic decreases are emerging for the four high-initial-information problems. Mean magnitude of change is also clearly age-related. The age correlation is −.09 ($p < .05$).

In the four low-initial-information problems, too, the monotonic decreases with age are seen at both times. Magnitude of change, however, is not related to age; the age correlation is zero.

The two-attribute problems with high initial information are major contributors to the relationship between age and change in total number of problems correct. It should be noted that these are not the most difficult problems. For every age group, problems with high initial information are correctly solved more frequently than those with low initial information. The differences are even more evident among the effectiveness measures (to be presented later); problems with high initial information are solved much more effectively than are problems with low initial information.

Estimates of Age Change. The fact that first-time data have been collected for 13 years makes it possible to calculate estimates of age changes within birth cohorts. This is conceptually similar to Schaie's cross-sequential analysis with independent samples except that to estimate age changes, within-cohort slopes are used rather than differences between means. Schaie samples each birth cohort at two (or more) points in calendar time. The two samples within a cohort are tested at different mean ages; that is, the second sample is measured seven years later than the first and is seven years older. In the BLSA, intake of new longitudinal participants occurs continually over time. Data are not collected during narrow time periods with long intervals between. Therefore, means at discrete times cannot be compared. Instead, for each birth cohort, first-time measures are plotted against calendar time. Declines with age within cohorts should show up as negative regressions of performance on time. Figure 1 is an example of a plot of number of problems correct for the cohort born from 1887 through 1896. Such slopes can be calculated for seven birth cohorts (there were only seven subjects in the cohort born prior to 1887 and, therefore, that cohort is not discussed) for

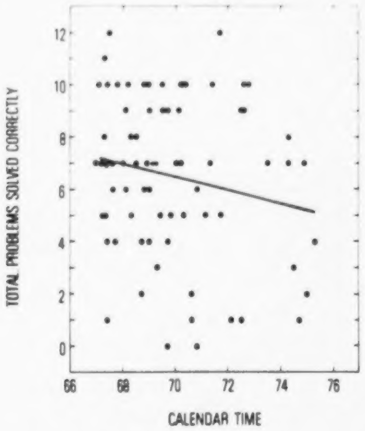


Fig. 1. Number of problems solved correctly in the first testing session for men born between 1887 and 1896. The abscissa is the date of testing.

number of problems correct. The results can be seen in Table 3. These slopes are near zero except for the cohort born from 1887 through 1896. That is the cohort plotted in Figure 1. Estimates of substantial age declines are showing up only for the oldest cohort, the men who were in their 70s at the beginning of the study.

These men can also be categorized into age groups to plot performance on calendar time. This is similar to Schaie's time-sequential analysis. If there is some systematic *nonmaturational* effect occurring during the 13 years of the study, it should show up as nonzero regressions within age groups as well as within birth cohorts. Table 4 shows the slopes for number of problems correct for the seven age groups in the cross-sectional analysis. All of these slopes are small and all but one are positive. The common age-group slope is .04, which is not statistically different from zero.

The slope of −.25 for the cohort born from 1887 through 1896 can be compared with zero and with the common age-group slope. Both comparisons

Table 3. Birth Cohort Slopes—Number of Problems Correct First Time

	N	Slope (per yr.)
1947–56	45	.12
1937–46	135	.09
1927–36	87	−.03
1917–26	156	−.09
1907–16	138	.02
1897–06	109	.00
1887–96	75	−.25
1877–86	7	—

Table 4. Age Group Slopes—Number of Problems Correct First Time

	N	Slope (per yr.)
20s	71	.08
30s	143	.06
40s	154	.03
50s	142	.05
60s	118	.03
70s	101	.05
80s	22	−.11
Common		.04

are statistically significant. The estimate of age change for the earliest-born cohort indicates a decline of one problem correct every four years. This is substantially larger than the mean longitudinal change of the oldest groups when measured directly (longitudinal repeated measures). Sequential analyses of independent samples avoid many of the attrition problems inherent in repeated measures (longitudinal) analyses. It is likely that the mean changes seen longitudinally are positively biased. In order to be included in the longitudinal analyses, the men not only had to continue in the study for at least six years after the first measures but had to solve all 12 problems both times. There is some evidence that these criteria are positively biasing, particularly for the older groups. In other words, the older participants who are included in the analyses are a highly select group, and their measures of change are likely to be underestimates of the actual changes (if everyone could be measured).

Effectiveness Measures. The problems were designed in such a way that each solution could be quantified in terms of information gain. For each selection (after the first) in every problem, the number of possibilities eliminated is determined as well as the number of possibilities that could be eliminated by an optimal selection at that point. These are converted to bits of information; then the ratio of the subject's bits and the optimal bits is calculated. These ratios range from zero (no additional information) to one (maximal information gain). The effectiveness measure for a subject is the mean of these ratios for all of his selections in a single problem. Within each problem, this measure is highly correlated with the number of selections required to solve that problem.

For each subject, effectiveness measures were averaged for problems which are logically identical: simple (one attribute), complex (high initial information), and complex (low initial information). Decade means of these averages for the simple problems are presented in Table 5 for the 391 subjects who had a valid average at both times of measurement (at least six years apart). These means decline monotonically with age when first measured and also when retested. Furthermore, the youngest group is improving and the men in their seventies

Table 5. Longitudinal Means—Effectiveness Measures, Simple Problems

	N	Mean 1st	Mean 2nd	Change
20s	10	.82	.87	.04
30s	35	.79	.79	−.01
40s	105	.77	.76	−.01
50s	105	.73	.71	−.02
60s	80	.71	.68	−.03
70s	49	.64	.58	−.06
80s	7	.58	.58	—

(when first measured) are showing the largest mean decline. The correlation of magnitude of change with age is −.14 ($p < .05$).

Similar results are emerging for the complex problems with high initial information (see Table 6). The age group means decline virtually monotonically at both times of measurement. Again, those in the youngest group are improving and the men in their seventies (initially) are showing the largest mean decline. The correlation between age and magnitude of change is −.12 ($p < .05$).

Although the effectiveness measures are substantially lower for the complex problems with low initial information, the pattern of results is somewhat similar to those of the other two types of problems (see Table 7). Age group means decline virtually monotonically at both times of measurement. The youngest group is improving; but, unlike the results of the other problem types, the men in their seventies (initially) are not declining substantially. For these problems the correlation between age and magnitude of change is only −.05.

These effectiveness data indicate substantial age differences cross-sectionally; and longitudinally, the youngest group is showing consistent improvement whereas the men in their seventies are declining. All correlations between age and change are negative, indicating that the magnitude of change in effectiveness is related to age with the oldest men tending to decline the most, quite similar to the results for number of problems correct.

One of the ways the effectiveness means are deflated is by overtly redundant

Table 6. Longitudinal Means—Effectiveness Measures, High Initial Information

	N	Mean 1st	Mean 2nd	Change
20s	10	.88	.91	.03
30s	32	.84	.85	.00
40s	94	.79	.79	−.01
50s	93	.76	.76	−.01
60s	72	.71	.72	.01
70s	37	.66	.58	−.09
80s	4	.75	.47	—

Table 7. Longitudinal Means—Effectiveness Measures, Low Initial Information

	N	Mean 1st	Mean 2nd	Change
20s	10	.66	.72	.06
30s	32	.66	.65	−.01
40s	99	.62	.64	.02
50s	86	.61	.59	−.02
60s	64	.57	.58	.01
70s	38	.52	.50	−.02
80s	3	.41	.54	—

selections, that is, repetitions of selections previously made within a problem. These, of course, are totally noninformative. Such repetitions occur mostly in the two-attribute problems with low initial information. They occur most frequently in problems solved incorrectly, less frequently in problems solved correctly by the elderly, and quite infrequently by the young. It should be noted that subjects know such repetitions are noninformative. Later in the same problem when subjects review their previous selections, they often discover the repetitions. When they do, they frequently attempt to cross them out on the paper, sometimes commenting with a four-letter expletive.

It appears that after many selections are made, the older subjects have substantial difficulty reviewing their selections and planning and carrying out their next selection. They seem to be suffering from information overload, and one way to deal with this is not to compare the tentative next selection (written but not yet verbalized to the experimenter) with all of the previous selections. When they make that comparison, they may lose sight of why they made the selection. As a result, they sometimes omit comparing their tentative next selections with their previous selections. Young subjects appear not to have that difficulty.

It seems therefore, that in complex problems, even when access to all current information is available for review (thereby reducing the memory load), the older men nevertheless experience information overload, and that contributes to their lower effectiveness (relative to younger men).

Summary of Results

Consistent cross-sectional age differences are found both in number of problems solved correctly and in measures of effectiveness. Furthermore, the direct measures of change (longitudinal, repeated measures) tend to be age-related; typically, the younger men improve and the older men decline. This is supported by the indirect estimates of change using regression analyses (sequential procedures) within birth cohorts and within age groups. Only the earliest born (oldest) birth cohort (in their 70s at the beginning of the study) are showing evidence of substantial decline with increasing age. These data indicate that in concept problems requiring reasoning, but with all current information available for review to reduce memory load, older men do not perform as well as younger men; and for many of the older men, performance declines in six or seven years. It is tempting to dismiss these longitudinal data because the mean declines are small and the correlations between age and magnitude of change are very small. These longitudinal findings, however, are almost surely conservative assessments of mean changes (due to the positive bias likely for the older, highly attrited groups). The substantial estimate of decline with age seen in the sequential analysis of the earliest born cohort is an indication that declines with age are larger than those found longitudinally for the highly select group of older men. Taken together, these preliminary analyses indicate that concept-problem-solving performance of even well-educated older men is not only poorer than that of young men, but declines with age even over a period of six or seven years.

Current Status and Future Directions

The literature on problem solving and aging was recently reviewed by Rabbitt (1977). I believe I am not distorting the tone of his chapter to describe the status at that time as disappointing. The domain of psychology labeled *problem solving* has a history of disarray. Not surprisingly, then, the area of problem solving in geropsychology is not well organized. What is surprising is the paucity of research. Rabbitt was struck by the contrast between the vast literature on problem solving in children and the scanty literature in aging.

Last year, my colleague, Leonard Giambra, surveyed the literature since Rabbitt's chapter and found about another dozen studies after 1974. One interesting aspect of the most recent studies is a focus on what is sometimes referred to as "intervention" studies. In these studies, experimental procedures are used to improve problem-solving performance of the elderly.

Some studies have shown that specific training improves performance of the elderly immediately after training. Of more practical significance, perhaps, are the results of a few recent studies which were designed to determine whether training has any lasting effects on problem-solving performance. Some training procedures were found to improve performance some time after training. Most intriguing was the finding by Labouvie-Vief and Gonda (1976) that the control group, which received no training but had practice equivalent to the trained groups, showed sustained effects (about two weeks later) equal in magnitude to that of the trained groups. These investigators interpreted this outcome as evidence that practice allows subjects to develop their own strategies, and that self-developed strategies are as likely to be retained as strategies provided in the training conditions. It is rather reassuring to me that the elderly can develop skills in solving unfamiliar problems without intervention when given opportunity to become experienced with the tasks.

One aspect of cognitive intervention research baffles me. Some investigators seem to assume that if performance of the elderly can be improved by training, then differences or changes with age in such performance are not attributable to maturation. Frankly, I do not understand this reasoning. If we provide a hearing aid to an older person whose hearing has declined with age, and that person can hear better when he uses the device, I doubt that we would say that this is evidence that the hearing impairment was not maturational. If we teach a mnemonic procedure to an older person whose memory has declined with age, and that person can remember better when he uses the mnemonic, why is that evidence that the memory impairment was not maturational?

One direction of research that Rabbitt advocated to account for age differences and changes in problem solving is to dissect complex performance into components. He suggested that contributions of memory, learning, information-processing rate, attention, and other cognitive components must be assessed to improve our understanding of problem solving and what happens to problem-solving performance with advancing age. I concur wholeheartedly.

Giambra has embarked on an ambitious enterprise in this direction. After many years of research in the mainstream of concept learning, Giambra concluded that group means based on one or a few problems per subject are not likely to enhance our knowledge about how an individual solves a complex problem. He decided that it is necessary to study individuals through many problems in order to understand how they go about solving. He has each subject solve more than 100 complex concept problems and attempts to model their performance. During some of the problems, subjects "think aloud" as they solve. The models are generated from the "thinking aloud" protocols and tested with the other problems. A successful model must predict performance throughout each problem.

The components of each person's model include some of the components mentioned by Rabbitt. Memory is very important. Each model includes parameters for how many instances are retained during the solution, which are retained, which aspects of instances are attended to, and how the information is organized. Giambra has succeeded in developing a model for a few subjects including a 96-year-old man. It will be extremely interesting if commonalities of the models can be found among problem solvers of the same age, of the same proficiency, or of the same combination of age and proficiency.

Recently, I had the opportunity to read a manuscript (Charness, 1981) describing an age study of rated tournament chess players in Canada. It was possible to obtain measures of components (such as memory) of the complex performances on chess problems. Even when the ratings of the players were statistically controlled, age was related to some memory measures accounting for performance on the chess problems. Age, however, was not related to performance on the chess problems, suggesting that a compensatory mechanism, such as more efficient search, allowed the older chess players to perform as well as the young. The memory decrement is rather surprising to me because I had a general impression that highly practiced memory skills were not affected by aging. When I played duplicate bridge, I frequently observed skilled players in their seventies who had no difficulty maintaining card counts during a hand and, at the end of a session, reconstructing virtually all the hands they had played. But now there is some evidence that highly practiced memory skills are age-related even among proficient performers.

So where do we stand and where are we going? At the descriptive level, it seems clear that there are age differences in problem-solving performance. Longitudinal data are scanty; and, based on what is available thus far, mean changes with age are small and tend to be found only late in life. It is likely, however, that these means are underestimates of changes with age due to the positive biases typically operating in longitudinal studies; and the effects of these biases are likely to be even larger in these problem-solving studies because only a select subgroup of returning subjects can be included.

These results are similar to many of the cross-sectional and longitudinal findings in studies of intelligence test performance. In Botwinick's recent review of intelligence and aging, he stated, "By and large, longitudinal studies show less decline than do cross-sectional ones; they may also show the decline starting later in life" (1977, p. 590). These results are rather different from the findings for memory-for-designs performance. In general, age differences in cross-sectional comparisons, longitudinal changes, and sequential analyses were quite consistent—small, if any, age effects in early adulthood, modest effects in middle age, and substantial age effects only late in life (Arenberg, 1978, 1982). Not surprisingly, different aspects of cognitive performance are showing somewhat different patterns of decline with age, but the finding common to virtually all aspects of cognitive performance is decline late in life.

There are indications of some progress, mostly not yet published, in analyzing components of complex problem-solving performance. Much more research of this type is needed. I would like to conclude on the optimistic note that, at least for some types of problems, when older people become familiar with them, performance improves without training, and the level of performance is maintained beyond the practice session. If that finding holds up generally, then some of the age differences in performance should be reduced by exposure to the problems.

ACKNOWLEDGMENTS

The author is indebted to Karen Douglas, Darrel Gray, Pat Hawthorne, Marian Hedrick, Barbara Hiscock, Steve Kanis, Joan King, Judy Plotz, and

Marcia Swartz for administering and scoring the data and to Judy Plotz for computer analysis of the data, as well.

References

Arenberg, D. Equivalence of information in concept identification. *Psychological Bulletin*, 1970, *74*, 355–361.

Arenberg, D. A longitudinal study of problem solving in adults. *Journal of Gerontology*, 1974, *29*, 650–658.

Arenberg, D. Differences and changes with age in the Benton Visual Retention Test. *Journal of Gerontology*, 1978, *33*, 534–540.

Arenberg, D. Estimates of age changes on the Benton Visual Retention Test. *Journal of Gerontology*, 1982, *37*, 87–90.

Botwinick, J. Intellectual abilities. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold, 1977.

Charness, N. Aging and skilled problem solving. *Journal of Experimental Psychology: General*, 1981, *110*, 21–38.

Denney, N. W. Aging and cognitive change. In B. B. Wolman & G. Stricker (Eds.), *Handbook of developmental psychology*. Englewood Cliffs, N.J.: Prentice-Hall, 1982.

Giambra, L. M., & Arenberg, D. Problem solving, concept learning, and aging. In L. W. Poon (Ed.), *Aging in the 1980s: Psychological issues*. Washington, D.C.: American Psychological Association, 1980.

Jerome, E. A. Decay of heuristic processes in the aged. In C. Tibbits & W. Donahue (Eds.), *Social and psychological aspects of aging*. New York: Columbia University Press, 1962.

John, E. R. Contributions to the study of the problem-solving process. *Psychological Monographs*, 1957, *71*, 1–39.

Labouvie-Vief, G., & Gonda, J. N. Cognitive strategy training and intellectual performance in the elderly. *Journal of Gerontology*, 1976, *31*, 327–332.

Rabbitt, P. Changes in problem-solving ability in older age. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold, 1977.

Schaie, K. W. Quasi-experimental research designs in the psychology of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold, 1977.

Stone, J. L., & Norris, A. H. Activities and attitudes of participants in the Baltimore Longitudinal Study. *Journal of Gerontology*, 1966, *21*, 575–580.

Wetherick, N. E. A comparison of the problem-solving ability of young, middle-aged and old subjects. *Gerontologia*, 1964, *9*, 164–178.

Wetherick, N. E. The inferential basis of concept attainment. *British Journal of Psychology*, 1966, *57*, 61–69.

Young, M. L. Problem-solving performance in two age groups. *Journal of Gerontology*, 1966, *21*, 505–509.

Birren, J.E.; Munnichs, J.M.A.; Thomae, H.; and Marois, M. (Eds.),
Aging: A Challenge to Science and Society. Vol. 3, Behavioural Sciences and Conclusions.
Oxford: Oxford U.P., 1983

Memory and Learning Do Decline Late in Life

David Arenberg

Introduction

I would like to present some research evidence which indicates diminished performance on learning and memory tasks late in life. These preliminary results are from the Baltimore Longitudinal Study and I will first explain why I believe they are important. A pervasive belief among people in the field of aging in the United States is that intellectual functioning does not decline even late in life except shortly before death. Much of this thinking is based primarily on longitudinal psychometric data, particularly the papers of Dr Warner Schaie and his colleagues. These data have been characterized by the term 'the myth of intellectual decline'. Recently, in response to a critical paper, Drs Baltes and Schaie clarified their use of the term 'myth' as meaning 'an uncritically accepted belief'. Unfortunately, many of us interpreted 'myth' to mean 'imaginary' or 'contrary to fact'. As a result, many gerontologists and people who work in the field of aging now believe that intellectual functioning is maintained even late in life.

Furthermore, it is easy to use 'cognitive functioning' for 'intellectual functioning', a short step from 'cognitive decline' instead of 'intellectual decline'. Then we have gerontologists, even some who are in cognitive gerontology, believing that *memory* and *learning* and *reasoning* do not decline late in life. It is an appealing belief, and it is understandable why it is so readily accepted, but I am convinced that it is wrong! Let me hasten to add that I am not referring only to speed or cognitive performance based upon speed; I am talking about *learning*, and *memory*, and *problem-solving*—aspects of behaviour we cherish and would like to maintain.

Evidence can be found for decline in performance late in life even in Dr Schaie's own data. Age gradients for the five subtests of the Primary Mental Abilities were reported by Schaie and Strother (1968a, b), Schaie, Labouie-Vief, and Buech (1973), and Schaie and Labouie-Vief (1974). One paper in 1968 described analyses of repeated measures at two points seven years apart. The other paper in 1968 compared independent samples of the same birth cohort measured at two different times. By comparing several birth cohorts in this way, estimates of age changes can be obtained without repeating the tests for any individual. In both papers, the means for the cohorts which were over 60 when initially measured declined in all five subtests.

Similarly, in 1973 and 1974, one paper described analyses of repeated measures of three-point data covering a 14-year period; and the other paper compared three independent samples of the same birth cohort measured at different times. Again, in both types of comparisons, repeated measures of the same individuals and independent samples born at the same time, the means declined in all five subtests for the cohorts which consisted of people over sixty when measured initially. For the most part, the mean declines over seven years and even over fourteen years were not large; but the consistency of these declines indicates to me that even psychometric performance declines late in life.

The investigation

The data I will report are from studies of verbal learning and memory, aspects of cognitive functioning not represented in the psychometric longitudinal data in the literature. The primary analyses were of cross-sectional data collected between 1960 and 1964 and of longitudinal data of the men who continued in the programme and were measured again six or more years after their first measure. We also compared independent samples of the same birth cohorts measured at different times. First measures obtained between 1968 and 1974 were compared with first measures obtained between 1960 and 1964 for six different birth cohorts. Although our sampling procedures were not nearly so elegant as those in Dr Schaie's studies, and the period within each time of measurement was several years, we decided to compare the independent samples with this idea in mind. If the age changes obtained in the repeated measures, the primary data, were supported by age differences within cohorts in the independent samples, we would be more confident that our longitudinal changes were not artefacts of attrition, repeated measurement, or other potential sources of bias.

The participants in the Baltimore Longitudinal Study are all males whose ages span the entire adult range with the vast majority between 30 and 80. They are, for the most part, healthy and educated and the group includes a high proportion of active or retired scientists, administrators, and professionals.

Paired-associate learning

We begin with paired-associate learning. In this task, each item consists of a stimulus component and a response component. The stimulus is the signal to say the response that has been paired with that stimulus. We used two consonants for each stimulus and a two-syllable word for each response, for example, TL-INSANE. Whenever the learner saw 'TL' he was supposed to say 'insane'. There were eight such items in each list.

Each participant was randomly assigned to one of two anticipation intervals. He was given 1.8 s (the short interval) or 3.7 s (the long interval) to respond to each stimulus. The dependent measure was the total number of errors to reach the criterion of one errorless trial.

The results at the short interval are shown in Fig. 1. The same format is used in most of the figures. The dependent measure is errors, so up is in the direction of poor performance. Age is the variable on the abscissa, but instead of age decades, the data were classified according to birth cohort. The youngest cohort was born between 1925 and 1932, the next cohort between 1917 and 1924, and so on, with the oldest cohort born between 1885 and 1892.

The cross-sectional means are shown as open circles connected by dotted lines. There were small age differences for the youngest groups, larger age differences in the middle groups, and a very large difference for the oldest groups. These data

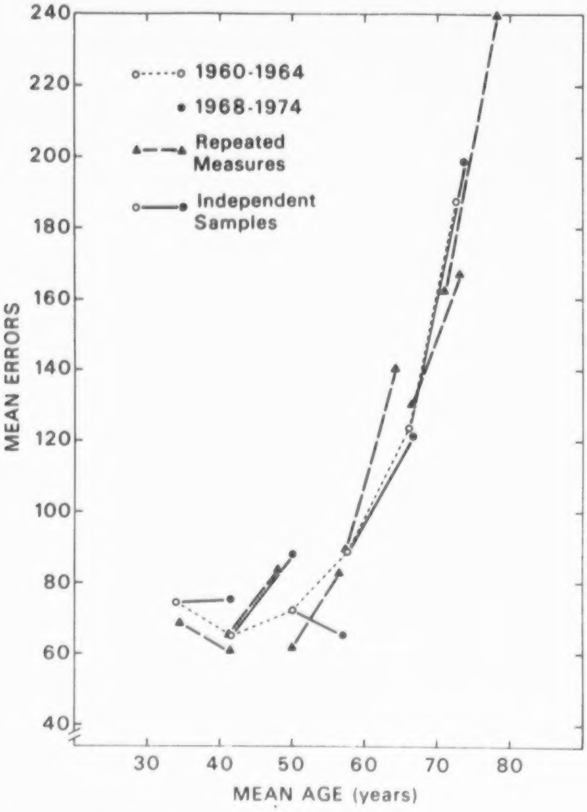


FIG. 1. Paired-associate learning—short interval.

are for all men born between 1885 and 1932 who were measured at the short interval between 1960 and 1964 ($N=153$).

Only some of those men continued in the programme and had a valid second measure at least six years later ($N=102$). Mean errors for those who returned are shown as filled triangles. The first and second measures are connected by dashed lines for each cohort. The youngest group improved slightly, the next two groups had a moderate increase in errors, the next two substantial increases in errors, and the oldest group had the largest mean increase in errors.

Ordinarily in a longitudinal study, we would stop at this point. However, as I mentioned earlier, we could compare performance of men born at the same time but measured at different times to determine whether these within-cohort comparisons supported the longitudinal results. Mean errors of men measured between 1968 and 1974 are shown as filled circles on the figure ($N=65$ for 5 cohorts). A solid line connects the two samples born at the same time, showing that for the latest born cohort, the younger subsample (measured early) was not different from the older subsample. The next two cohorts showed small differences in different directions. But the two oldest cohorts showed large differences between the younger and older men born at the same time.

The longitudinal results based on repeated measures were well supported by the comparisons of independent samples born at the same time. All of these data indicate substantial declines in verbal learning late in life.

The other verbal learning data are not so compelling, but the same story emerges nevertheless. In Fig. 2 we see the same kind of data for men who learned at a longer anticipation interval, i.e., they had more time to respond to each stimulus. The cross-sectional data ($N = 165$), shown as open circles connected by a dotted line, resulted in little age difference until the 60s. The older men benefited a great deal from having the additional time to respond. The longitudinal data ($N = 111$), shown as filled triangles connected by dashed lines, resulted in increases in errors for all cohorts with the two oldest groups showing the largest increases (declines in performance).

Again independent samples within a cohort are shown as an open circle connected to a filled circle by a solid line ($N = 50$ for 5 cohorts in later sample). All

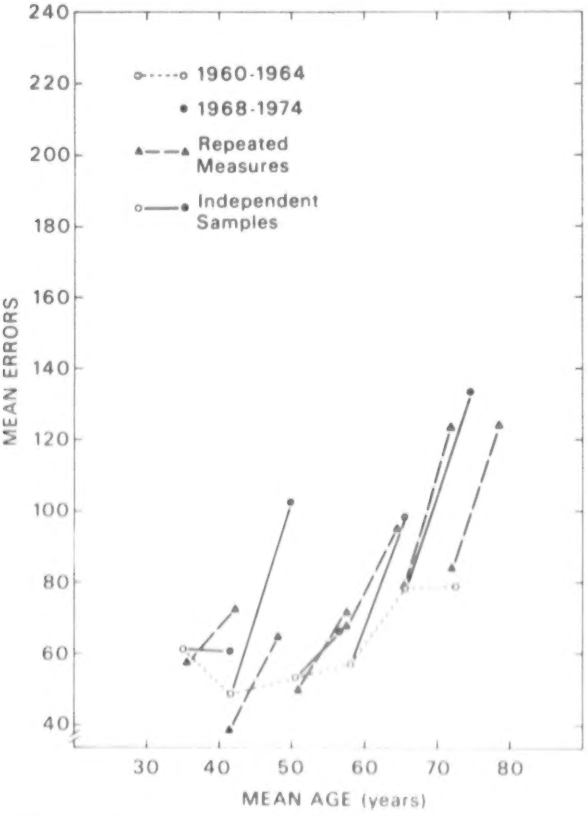


FIG. 2. Paired-associate learning—long interval.

but the youngest cohort showed increases in errors and, except for the second cohort, the largest increases were found for the oldest cohorts. Again the comparisons of independent samples within cohorts support the longitudinal findings of declines late in life even at a pace the older men could handle quite well when measured initially.

Serial learning

Now we move on to the serial-learning study. The task was to learn a list of twelve words in serial order. The list was presented in the same order repeatedly. Each word, when presented, was the signal to say the next word in the list. Participants were given 3.8 s (short interval) or 5.6 s (long interval) to respond to each signal. As in paired-associate learning, the dependent measure was the number of errors to reach one errorless trial.

The cross-sectional results at the short anticipation interval ($N = 157$) were similar to the paired-associate results. In Fig. 3 the open circles connected by a

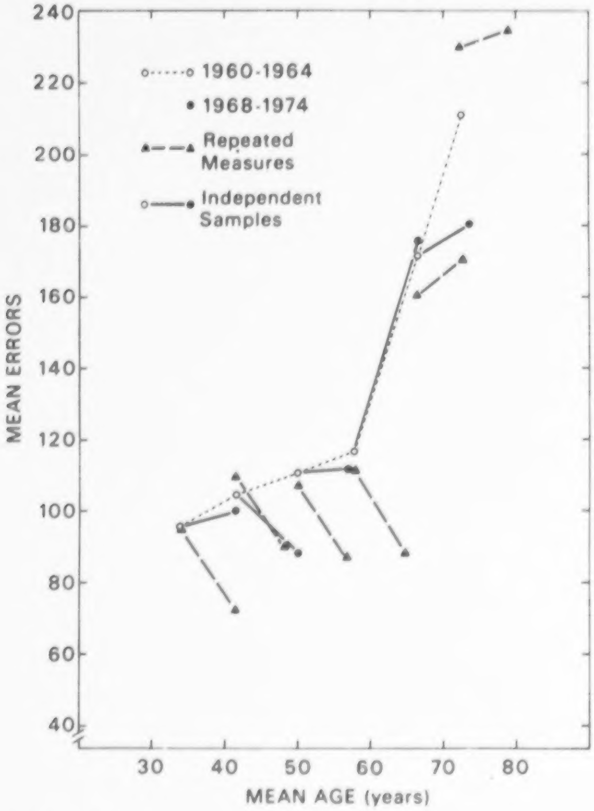


FIG. 3. Serial learning—short interval.

dotted line indicate small age differences for the younger groups and large age differences for the oldest groups.

The longitudinal results ($N = 104$) were quite different, but the relative pattern was similar. These results are shown by filled triangles connected by dashed lines. The youngest age groups decreased in mean errors. Although the words in the list used for the second measure were structurally similar (in form and frequency) to the words in the first list, we have some external evidence that the second list was easier. That would account for some of the improvement for the younger groups. Despite the easier second list, the two oldest groups showed increases in mean errors, i.e, their performance declined.

The comparisons of independent samples within cohorts is not affected by list differences. Everyone learns the same list. These cohort differences are shown by an open and a filled circle connected by a solid line ($N = 68$ for 5 cohorts in later sample). Although the results are not systematic, again the largest differences between the younger and older subsamples were found for the oldest cohorts.

The cross-sectional results ($N = 158$) for the longer interval were also similar to the paired-associate data. Fig. 4 shows that the means (open circles connected by a dotted line) did not increase monotonically with age, but there was a trend of increasing errors with increasing age.

The longitudinal data (filled triangles connected by dashed lines, $N = 102$) show small and inconsistent mean changes except for the oldest group for whom errors increased greatly.

The comparisons of independent samples within cohorts (an open circle connected to a filled circle by a solid line) were not affected by list differences, and these results showed increases for all cohorts ($N = 51$ for 5 cohorts in later sample) with the largest increase found for the oldest cohort.

The results for serial learning were consistent with those for paired-associate learning. These findings indicate decrements in verbal-learning performance late in life.

When the two learning tasks were given for the first time during the period from 1964 to 1968, they were administered self-paced; that is, whenever the learner was supposed to respond, he had as much time as he needed to say the word. These samples were smaller than the paced samples, and no later sample was available to carry out sequential analyses. Although the mean age changes were not substantial, the relationship of age to change in performance was statistically significant for both serial and paired-associate learning so that even when subjects were given as much time to respond as they needed, a condition that reduces cross-sectional age differences, the oldest subjects declined the most in both tasks.

Memory for designs

We also studied memory for designs using the Benton Revised Visual Retention Test which is predominantly non-verbal. Each of the ten designs is shown for ten seconds and then is removed. The task is to reproduce each design from memory immediately after inspection of that design; no time limit is

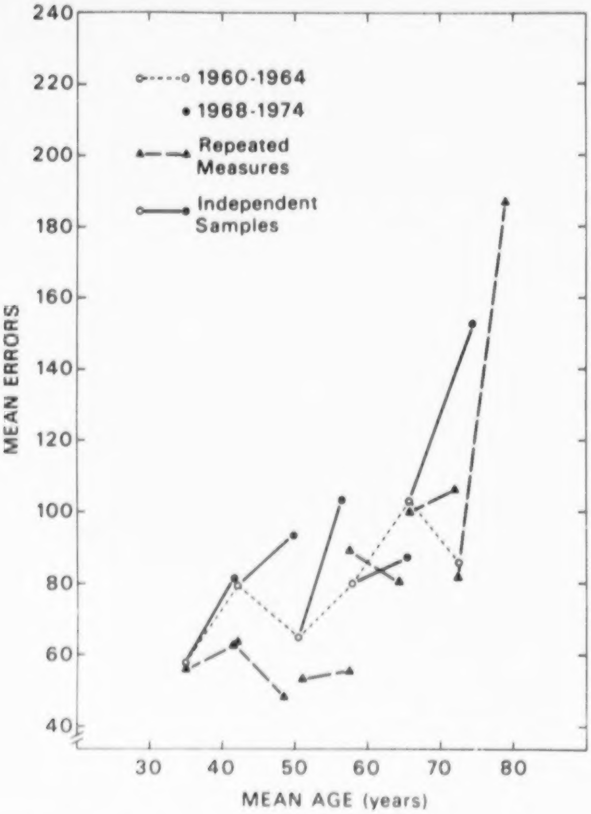


FIG. 4. Serial learning—long interval.

imposed. Each design consists of one or three geometric figures. The designs were scored independently by two psychologists in accordance with the manual, and the few disagreements were resolved by them or with a third scorer, if necessary. The dependent measure was the number of errors in all ten designs.

The cross-sectional, longitudinal, and sequential analyses are shown in Fig. 5; the format is the same as the one used for the verbal learning figures.

The cross-sectional results are shown as open circles connected by a dotted line ($N = 382$ for 6 cohorts born between 1885 and 1932 in early sample). Small differences were found at the younger ages, but substantial increases in errors were found for the oldest cohorts.

The longitudinal results are shown by filled triangles connected by dashed lines ($N = 263$ for 6 cohorts). The changes follow closely the cross-sectional age curve. The young groups changed very little, the middle groups declined moderately, and the oldest group declined substantially.

Comparisons of independent samples from the same birth cohorts are shown by an open circle connected to a filled circle by a solid line ($N = 198$ for 6 cohorts

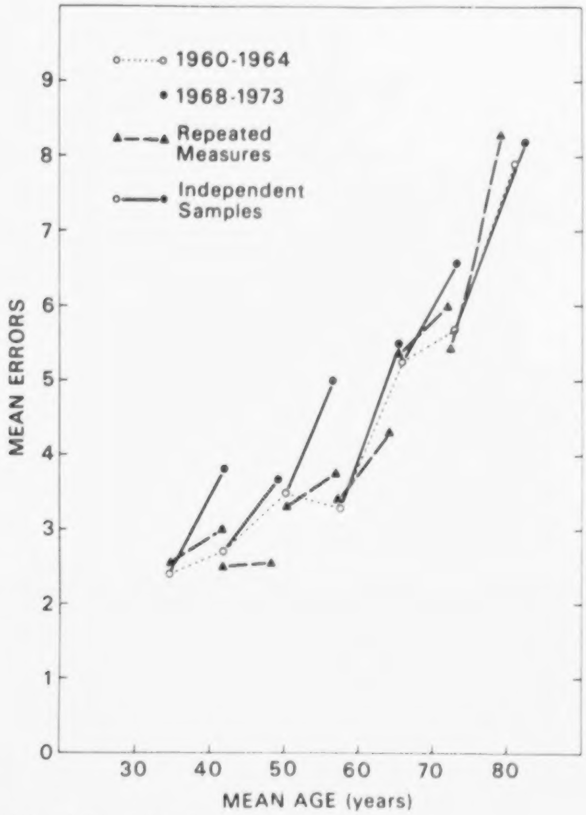


FIG. 5. Benton Visual Retention Test—errors.

in later sample). These estimates of change were all in the direction of decline, with the largest decline in the oldest cohort.

The cross-sectional, longitudinal, and sequential changes and differences all point to substantial declines in non-verbal memory late in life.

Thus far we have been looking only at means. Fig. 6 shows a cumulative frequency curve of changes for each decade group to provide some indication of the distribution of longitudinal changes. Each point indicates the proportion of subjects in an age group who increased a specific number of errors or less. For example, the second point on the abscissa represents no change. We can see that 61 per cent of the men below 40 performed as well (or better) the second time as the first. More than half of the men in their 40s, 50s, and 60s did not decline; but only 28 per cent of the men in their 70s when first tested performed at least as well the second time as the first.

If we move across to an increase of three errors, we see that more than 90 per cent of the men in the two youngest groups changed no more than three errors. More than 80 per cent of the men in their 50s and 60s initially performed that well.

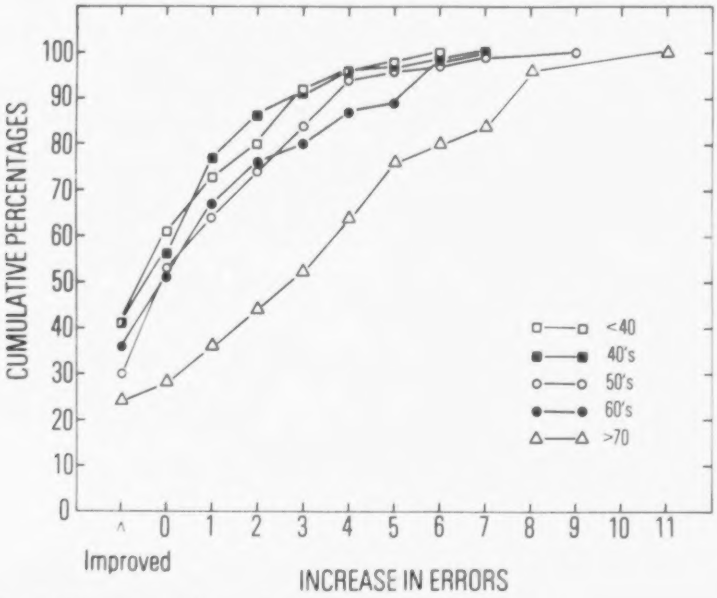


FIG. 6. Cumulative proportions—Benton change.

But only 52 per cent of the men initially in their 70s declined by no more than three errors; almost half of the oldest group declined more than that. And one-fourth of them declined by more than five errors; only 11 per cent of the men in their 60s and very few of the younger men declined by that much.

Fig. 6 shows that the large mean increase in errors for the men in their 70s was not attributable to a few men who changed drastically. Rather what we see is a shift of the entire distribution in the direction of larger decrements in performance.

We have replication data for this non-verbal memory test. Analyses based on a smaller sample of men whose first Benton performance was during the period from 1965 to 1968 closely replicated the results of the earlier data. Errors increased with age cross-sectionally. Mean changes were in the direction of improvement for a very small sample below 40 initially, no change for the 40s, small declines for the 50s and 60s, and substantial declines for the men 70 or older when first tested.

All the data for the Benton test of non-verbal memory tell a consistent story. Declines in the early and middle adult years are small, but late-in-life performance declines substantially. Note that in this task, subjects have as much time as they need to reproduce the designs. The evidence is quite clear. For this task, memory declines late in life.

Conclusions

A central topic of this book is how further research might improve the quality of life in old age. First, I believe we must dispel the misconceptions that learning and memory performance are maintained late in life. Only if we accept that a problem exists can we do something about it. The results summarized in this presentation indicate that learning and memory do decline late in life even for educated, relatively healthy men. A problem clearly exists. What needs to be done?

One research approach is to characterize those old people who do not decline. It should be pointed out that I have been talking mostly about means; but in every age group, even the oldest, we found some individuals whose performance did not decline and was indistinguishable from that of young adults. Even in the memory-for-designs task, 28 per cent of the men over 70 when tested initially showed no decline whatsoever. What is it about those old individuals that accounts for maintenance of their performance levels? The more we know about levels and changes in other important variables, the more likely we will be able to answer questions about why some people decline and others do not.

Another important area of research to improve the quality of life in old age is to devise mnemonics and other training procedures specifically for the aged to minimize deficits in learning and memory. For example, in our laboratory, the method of loci, a mnemonic procedure, was modified to provide an effective aid to learning items in a list. An older person can learn lists of words, for example a shopping list, quite readily by applying a well-established mental trip through his residence visualizing an item at each (predetermined) stopping place. At recall, the mental trip is repeated, and the stopping places serve as effective retrieval cues for remembering the words. This procedure capitalizes on the highly overlearned spatial lay-out of a person's residence. The method is particularly well adapted for the old because it does not require learning an elaborate new memory scheme which may create an information overload so typically experienced by older learners.

One of the most frequent complaints of the elderly is their difficulty remembering names that go with familiar faces. If a mnemonic procedure could be developed for the old to improve recall of names, it would be a substantial contribution to the quality of life in old age.

References

Arenberg, D. (1967). *J. Gerontol.*, **22**, 411.

—— (1978). *J. Gerontol.*, **33**, 534–40.

Baltes, P. B. and Schaie, K. W. (1976). *Amer. Psychol.*, **31**, 720.

Benton, A. L. (1963). *The revised visual retention test: clinical and experimental applications* (3rd edn). Psychological Corp., New York.

Robertson-Tchabo, E. A., Hausman, C. P., and Arenberg, D. (1976). *Educ. Gerontol.*, **1**, 215.

——, Arenberg, D., and Costa, P. T. Jr. (1979). In *Brain function in old age: evaluation of changes and disorders* (Bayer Symposium VII), p. 151. Springer Verlag.

Schaie, K. W. (1974). *Amer. Psychol.*, **29**, 802.

—— and Labouvie-Vief, G. (1974). *Develop. Psychol.*, **10**, 305.

——, ——, and Buech, B. U. (1973). *Develop. Psychol.*, **9**, 151.
—— and Strother, C. R. (1968a). *Multivariate Behav. Res.*, **3**, 259.
—— and —— (1968b). *Psychol. Bull.*, **70**, 671.

Human Biology, September 1977, Vol. 49, No. 3, pp. 495-514.

Fat Redistribution and The Changing Body Dimensions of The Adult Male

By Gary A. Borkan[1] and Arthur H. Norris[2]

ABSTRACT

This study was undertaken to demonstrate the role of subcutaneous fat in altering external body dimensions in adult males. The sample population is composed of U.S. males between the ages of 25 and 84 years. Both cross-sectional and longitudinal data are available for these individuals. Weight of fat in the cross-sectional sample is found to be relatively constant with age but fat-free weight declines markedly. Data from soft-tissue radiographs reveal that in the trunk, subcutaneous fat increases in the region of the greater trochanter but decreases in the abdominal region through middle age. Abdominal diameter increases during this period, however, indicating enlargement or sagging of the abdominal contents. In the extremities, diameter of the calf and arm decline while fat is relatively stable, indicating loss of lean tissue with age. The present study agrees generally with earlier findings that age changes in body dimensions leading to thin extremities and thicker trunk are only partly attributable to fat redistribution.

The size, shape, and proportions of the human body are not stable during the long period of adulthood. This may be demonstrated by casual observation of individuals as they age, and has also been indicated in cross-sectional population samples (Hooten and Dupertuis, 1951; Stoudt et al. 1965, 1970; Howells and Bliebtreu, 1970; Damon et al. 1972). Research on adult body composition reveals that all the major tissues of the body undergo considerable alteration with age. Lifelong changes in the dimensions of the skeleton have clearly been demonstrated (Garn, 1970). A decline in lean body mass during adulthood has been shown using a variety of techniques (Brožek, 1952; Norris et al. 1963; Oleson, 1965; Forbes, 1976). The proportion of fat in the body has been found to be higher in older age groups in a number of studies (Brožek, 1952; Škerlj et al. 1953; Friis-Hansen, 1965; Krzywicki et al. 1967). Change in these body tissues during aging not only affects external size, but also alters

[1] Center for Human Growth and Development and Department of Anthropology, The University of Michigan, Ann Arbor, Michigan 48109 USA.

[2] Clinical Physiology Branch, Gerontology Research Center, National Institute on Aging, National Institutes of Health, U.S. Department of Health, Education, and Welfare, Bethesda, and Baltimore City Hospitals, Baltimore, Maryland 21224 USA.

body proportions. This occurs because the relationship between the amount of fat, lean tissue, and bone in various parts of the body is variable in an individual. This effect is most clearly evident in subcutaneous fat distribution and its changes with age.

In 1952, Brožek demonstrated the truth of the adage "adulthood is the period when growth stops except in the middle." Measuring a large number of skinfold sites in a population of adult males, he showed that fat in the trunk increases with each succeeding age group but fat in the distal extremities does not, and may even decrease. This phenomenon was also demonstrated in a similar sample of adult females by Škerlj et al. (1953, 1959). On the basis of calculated subcutaneous fat weight and total fat weight, these authors showed that internal fat increases relatively more than subcutaneous fat during adulthood, and thus increases its proportionate contribution to total body fat. Garn and Young (1956) showed in a cross-sectional sample that fat is lost on the anterior part of the lower leg with each succeeding age group, but is gained in the lateral part of the upper thigh. Such findings have given rise to the concept of fat redistribution (or rearrangement) in adulthood. Data providing evidence of differences in fat distribution with age have appeared in recent cross-sectional anthropometric surveys as well (Glanville and Geerdink, 1970; Damon et al. 1972; Montoye et al. 1975).

The present study analyzes changes in subcutaneous fat distribution in adult males with respect to changes in other body compartments. Through a comparison of body part diameter and subcutaneous fat thickness, the relative roles of fat versus lean tissue and bone in altering external dimensions are revealed. Fatness at specific sites is also compared with whole body compartments evaluated biochemically. Furthermore, the use of partly longitudinal data in this study serves to establish whether trends observed in cross-sectional sampling actually reflect changes which occur in the individual.

MATERIALS AND METHODS

Sample

The sample consists of 699 Baltimore-Washington area males, aged 20 to 92 years, who have participated in the Baltimore Longitudinal Study of the Gerontology Research Center, National Institute on Aging, Baltimore, Maryland. The vast majority are white (97%), middle class, and hold college or graduate degrees. Social and demographic characteristics of this group of men have been described by Stone and Norris (1966).

Radiographic Data

A total of 971 soft-tissue radiographs of 699 individuals, taken between 1958 and 1973, were available for analysis. Radiographic procedures and techniques were identical to those described and illustrated by Garn (1954). Each 7 × 17 inch film contains views of seven body sites on the trunk and limbs. Measurements were made of skin and fat combined because differentiation of these two tissues is difficult radiographically. These measurements have traditionally been referred to simply as "fat measurements." A Helios dial-reading caliper calibrated to .05 mm was used throughout the study. For a comparison of the radiographic and skinfold methods of assessing fatness, see Garn and Gorman (1956).

Sites of fat measurements on the trunk were bony landmarks such as the top of the greater trochanter; and for the calf and forearm, the widest part of the limb was used. Fat measurements were attempted at the following locations on each film: 1. anterior calf, 2. posterior calf, 3. medial calf, 4. lateral calf, 5. lateral to greater trochanter, 6. lateral to top of greater trochanter, 7. lateral to anterior-superior spine of iliac crest, 8. lateral to top of iliac crest, 9. lower part of thorax (lowest rib), 10. medial arm and, 11. lateral arm. A number of sites were not measurable on a majority of the films, usually because of improper body positioning or film handling. Reproducibility of measurements was tested by remeasuring a sample of 20 films at the end of the study. Correlations between the two sets of paired measurements averaged greater than .95.

Additional Data

Additional information on each individual was obtained from an anthropometric and body composition survey of the subjects in the Baltimore Longitudinal Study. These data include anthropometric body circumferences at sites corresponding to each radiographic fat measurement, height, weight, and age. Biochemical assessment of extracellular water (determined by sodium thiocyanate dilution), and total body water (determined by antipyrine dilution) were also obtained.

From these data a number of new variables were calculated. To aid comparison with the subcutaneous fat data the body circumferences were converted to diameters based on the assumption that the body part is circular. Using these values, muscle plus bone areas were calculated for the arm and leg. From the data on total body water and body weight, the percentage of body water was determined. The percentage of fat was calculated from the equation (Moore et al. 1963):

$$\%\text{body fat} = 100-\left(\frac{\%\text{body water}}{.732}\right)$$

and from this result;

$$\text{total body fat (kg)} = \%\text{body fat} \times \frac{\text{body weight (kg)}}{100}.$$

Estimates of fat-free percentage of the body and fat-free weight were determined on the basis of these values. Extracellular water was subtracted from total body water to obtain the percentage of intracellular water.

Data Analysis

The 971 x-ray films measured in this study represent a total of 470 individuals with a single film, 209 with two, and 25 with three or more. The differences in number of films per individual are the result of change in Baltimore Longitudinal Study policy regarding repeat soft-tissue radiographs. Statistical tests (Student's t) were made to determine whether the individuals with two or more films comprise a different subpopulation than the group with only one. It was found that both in physical and demographic terms, there is no difference between these two groups.

Two separate data sets were used in the analysis. A cross-sectional sample consists of first visit data for 699 individuals. This sample was divided into ten-year age categories beginning with age 25 and ending at 84. A longitudinal sample of 234 individuals with two x-ray films was also analyzed. This data set contains only the first two radiographs for individuals with three or more. The interval between the first and second films averaged 6.8 years. For each individual in the longitudinal sample, increment per year values were calculated for all variables. Each individual was placed in one of the ten-year age categories described above based on the average age between the two test visits. Third and fourth radiographs were available for very few individuals and were not included in the analysis.

RESULTS

Cross-sectional Sample

Median values by age for weight, height, and the various indicators of body composition are given in Table 1. Both weight and height decline steadily with each age group in this sample. With respect to various body compartments, total body water as percent of body weight declines in younger adults and later is rather constant. The proportion of fat relative to body weight increases with age. Since body weight declines with age in this sample, calculated fat weight changes relatively little through the six age groups. Thus, calculated fat-free weight declines with age. In addition, extracellular water percentage increases with age, whereas intracellular water decreases. The decline in intracellular water percentage also indicates a decline in lean tissue percentage with age.

The data on the various body sites also gives evidence of a decline in fat-free tissue with age. In the forearm, diameter is constant between the ages of 30 and 60 but declines markedly thereafter (Table 2, Fig. 1). Fat on the medial side of the forearm increases during adulthood and lateral arm fat first increases and then decreases slightly. Calculated arm muscle and bone areas show a steady decline during this period.

Calculated abdominal diameter increases considerably with age (Table 2, Fig. 2). Fat at the lower thoracic, iliac crest, and iliac spine sites all decline in thickness. This indicates expansion or sagging of internal abdominal contents.

In the lower trunk (Table 2, Fig. 3) the diameter at the greater trochanter increases after an initial decline, peaks at age 60 and decreases thereafter. Fat lateral to the top of the greater trochanter parallels this same trend. Fat lateral to the middle of the greater trochanter increases and then decreases in a similar, but not as well defined, manner.

The diameter of the calf (Table 2, Fig. 4) declines throughout adulthood. Of the four calf fat measurements only anterior calf fat shows a similar decline. Fat lateral, medial, and posterior to the calf decreases through middle age and then increases. Calculated calf muscle and bone area declines rather steadily in adulthood, its pattern being similar to that of the arm.

Intercorrelations between the fat measurements were determined from the entire cross-sectional sample using data of all ages combined (Table 3, upper portion). Mean intercorrelations of each fat measurement

Table 1

Body Composition Medians by Decade

	Age Group											
	25-34		35-44		45-54		55-64		65-74		75-84	
	N	Mdn	N	Mdn	N	Mdn	N	Mdn	N	Mdn	N	Mdn
Age (yrs)	76	31.7	152	40.5	183	49.6	129	59.4	116	70.1	29	77.6
Weight (kg)	76	81.6	152	78.7	183	77.1	129	78.0	116	76.5	29	73.1
Height (cm)	76	180.7	152	177.5	183	176.2	129	175.8	116	174.3	29	174.8
Total Body H_2O %	53	54.8	105	53.9	134	52.6	102	53.3	85	52.8	19	53.6
Total Body Fat %	53	25.1	105	26.4	134	27.5	102	26.8	85	27.9	19	26.7
Fat Weight (kg)	53	19.0	105	20.2	134	20.9	102	19.9	85	20.2	19	20.4
Fat-free Weight (kg)	53	59.7	105	56.8	134	56.5	102	57.7	85	54.6	19	52.3
Extracellular H_2O %	56	23.0	112	23.2	139	23.2	110	23.7	91	24.8	19	25.7
Intracellular H_2O %	53	31.3	104	31.0	133	29.8	102	29.4	19	25.7	18	25.5

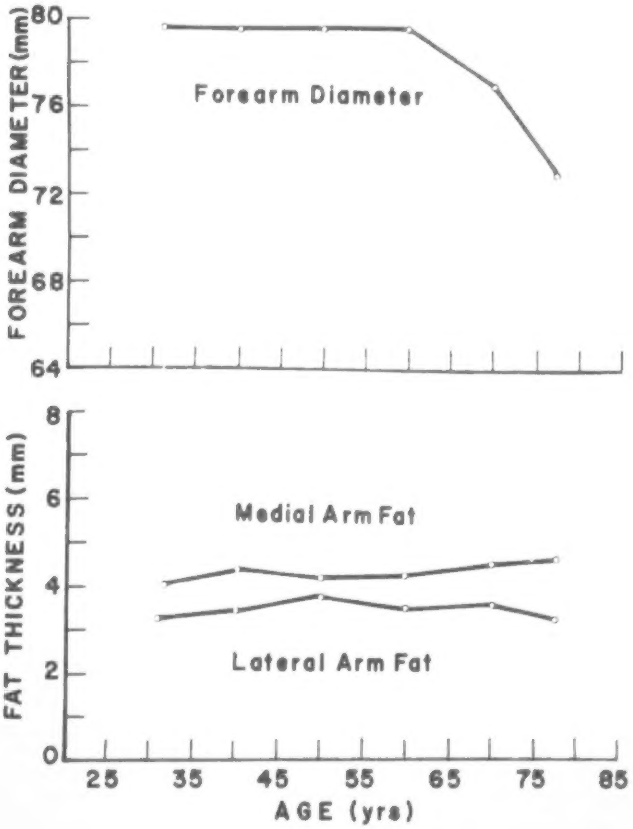


FIG. 1. Age trends in subcutaneous fat thickness and limb diameter at the widest part of the forearm. In the older ages there is a marked decline in diameter which is not attributable to change in fat thickness.

with the ten other fat measurements are also given in Table 3 (correlation values were first z-transformed and the mean of the z-transforms was then reconverted to r). Trunk fatness measurements have the best mean correlations overall. Mean correlations for medial and lateral sides of both leg and arm are all similar and are higher than the mean correlations for the anterior and posterior sides of the leg.

The lower part of Table 3 contains the correlation coefficients of each fat measurement with body weight and with fat weight. Fat thickness measurements of the trunk have the highest correlations with body weight. Specifically, measurements at the ilium and lower thorax appear to be the best predictors of body weight, and medial arm is the best predictor among the extremity sites. That correlations with fat weight are

Table 2

Body Composition Medians by Decade, Grouped by Site

	Age Group											
	25-34		35-44		45-54		55-64		65-74		75-84	
	N	Mdn	N	Mdn	N	Mdn	N	Mdn	N	Mdn	N	Mdn
Forearm												
Medial Fat	43	4.1	105	4.4	100	4.2	67	4.3	59	4.6	18	4.7
Lateral Fat	75	3.3	143	3.5	170	3.8	114	3.5	106	3.6	25	3.3
Muscle Area (cm^2)	34	44.2	80	40.9	74	40.1	49	38.0	43	37.1	12	35.4
Diameter	56	79.6	113	79.6	134	79.4	102	79.6	85	77.0	23	73.2
Middle Trunk												
Lower Thoracic Fat	58	15.0	109	13.9	135	13.3	95	13.4	91	12.1	23	12.2
Iliac Crest Fat	38	20.1	83	17.8	97	18.1	64	17.6	73	13.8	17	13.9
Iliac Spine Fat	25	31.1	62	28.4	78	26.4	45	28.6	53	22.9	15	24.4
Iliac Level Diameter	55	270.6	113	269.0	131	272.2	102	283.4	84	285.0	23	293.0

Table 2

Body Composition Medians by Decade, Grouped by Site

	Age Group											
	25-34		35-44		45-54		55-64		65-74		75-84	
	N	Mdn	N	Mdn	N	Mdn	N	Mdn	N	Mdn	N	Mdn
					Lower Trunk							
High Trochanter Fat	19	15.1	23	16.2	47	16.8	41	17.4	38	14.5	8	14.1
Lateral Trochanter Fat	21	18.1	34	18.9	34	16.7	35	17.3	32	18.5	5	15.6
Trochanter Level Diameter	56	307.2	113	302.4	131	302.4	102	308.8	84	305.6	23	300.8
					Calf							
Anterior Fat	75	3.2	146	2.9	171	2.8	117	2.6	98	2.4	24	2.4
Lateral Fat	71	5.2	136	4.9	162	4.8	103	4.7	86	4.8	21	5.2
Posterior Fat	28	8.1	57	6.4	61	6.0	54	5.7	37	6.2	12	6.7
Medial Fat[1]	53	6.2	113	6.4	137	5.8	101	6.2	77	6.3	22	6.3
Muscle Area (cm^2)	53	79.8	105	78.4	123	80.1	95	77.6	76	74.0	19	70.8
Diameter	56	113.0	112	113.0	133	111.4	102	111.4	85	109.8	23	108.2

[1] Unless otherwise noted, all units are millimeters.

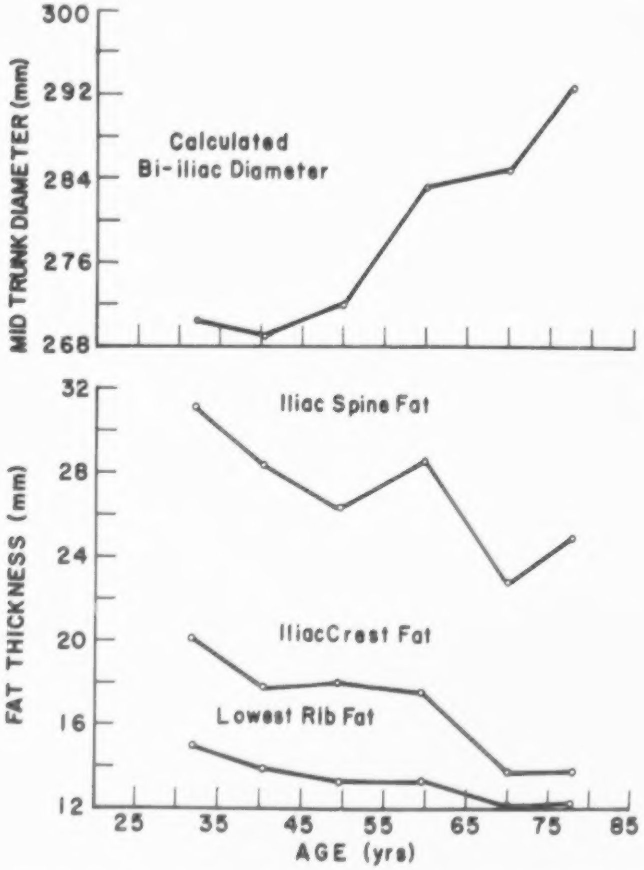


FIG. 2. Age trends in subcutaneous fat thickness and trunk diameter (calculated from circumferences) at the level of the iliac crest. Increase in middle trunk diameter accompanies loss of fat subcutaneously in this region. These opposing trends indicate expansion or sagging of abdominal contents, as well as a possible increase in internal fat.

lower than those for body weight reflects the higher standard error of the total body water determinations used to calculate fat weight.

The cross-sectional sample was divided into three age groups (25-44, 45-64, and 65-84) and in each group fat values were correlated with total weight and fat weight (Table 4). Mean z-transformations of the correlations were calculated (and reconverted to r) for three body regions (trunk, forearm, and calf). Correlations of subcutaneous fat with weight show that in all age groups the trunk sites have the highest values. Correlations with fat weight show that the trunk measurements remain good indicators of

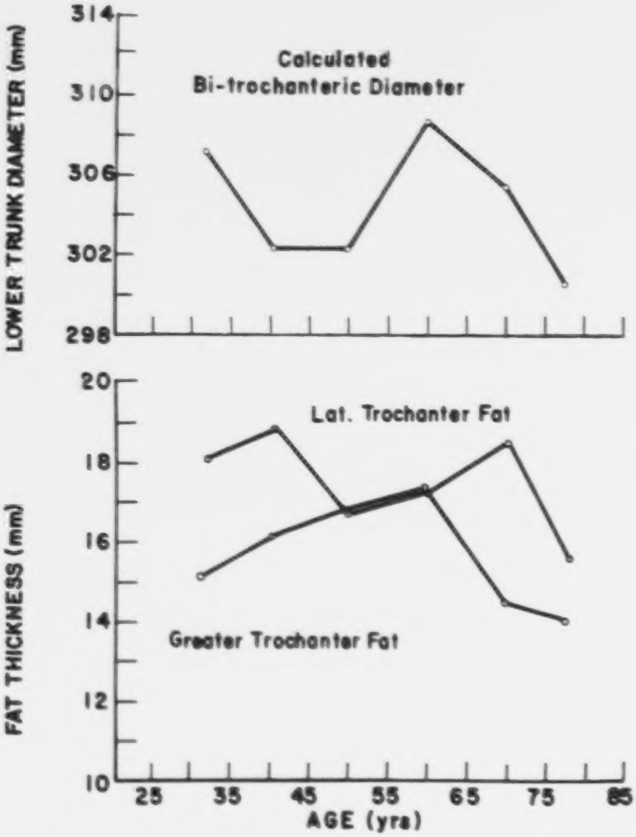


FIG. 3. Age trends in subcutaneous fat thickness and calculated trunk diameter at the level of the greater trochanter of the femur. Diameter and fat thickness change similarly with age, tending to increase through middle age and then decrease. Change in diameter appears to be largely accounted for by changes in fat thickness at this site.

body fat in all age groups although there is a decline in correlation values with age for the forearm and calf.

Longitudinal Sample

Yearly increments are reported for variables with a total N of greater than 35. Small sample sizes were the result of considerable missing data in the radiographic fat measurements, and body circumference increments were not available at all because these measurements were made only at the first visit of each subject. The longitudinal data are thus included

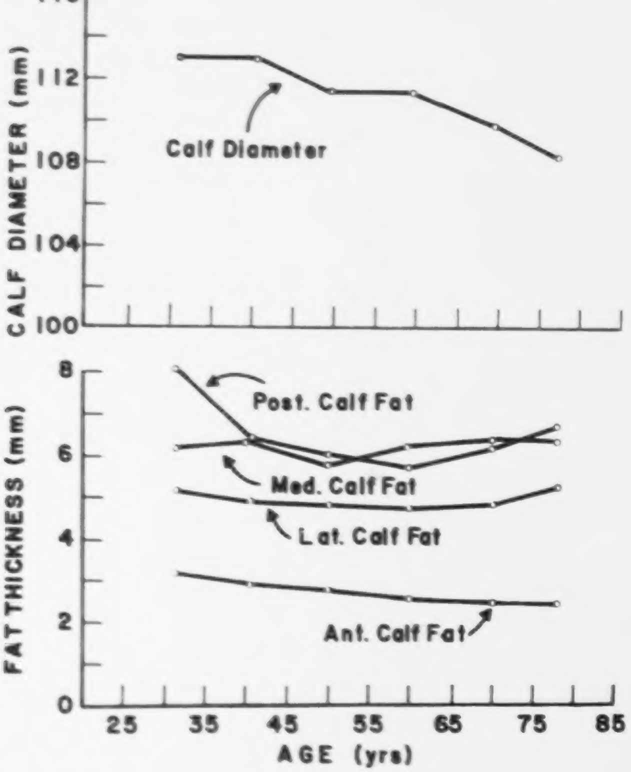


FIG. 4. Age trends in subcutaneous fat thickness and limb diameter at the widest part of the calf. Similarities with the forearm (Fig. 1) are indicated, with decline in limb diameter unaccompanied by loss of fat.

primarily for purposes of comparison with the larger cross-sectional sample (Table 5).

Viewed longitudinally, weight increases in the first three age groups and declines during the period between 55 and 74. The oldest individuals show a marked increase in weight and this seems to alter an apparent trend. This may be due to differential survivorship because the very old group may be composed of unusually healthy individuals. Height decreases in each age group to a greater and greater degree. Among measures of lean tissue, calculated fat-free weight declines in all age groups except the last. Body fat increments are highly positive in the first age group and decrease gradually, and the oldest group has a markedly negative increment.

Table 3

Intercorrelations between Fat Measures, and between Fat Measures and Body Weight and Total Fat Weight

	Ant. Calf	Post. Calf	Med. Calf	Lat. Calf	High Troc.	Lat. Troc.	Iliac Crest	Iliac Spine	Low. Thor.	Lat. Arm	Med. Arm
Anterior Calf		.41	.52	.55	.17	.23	.21	.29	.31	.26	.43
Posterior Calf	.41		.48	.55	.11	.14	.10	.15	.21	.27	.31
Medial Calf	.52	.48		.67	.43	.44	.32	.35	.43	.39	.46
Lateral Calf	.55	.55	.67		.30	.46	.22	.26	.37	.39	.49
High Trochanter	.17	.11	.43	.30		.93	.67	.61	.53	.43	.62
Lateral Trochanter	.23	.14	.44	.46	.93		.50	.70	.60	.34	.51
Iliac Crest	.21	.10	.32	.22	.67	.50		.84	.77	.42	.40
Iliac Spine	.29	.15	.35	.26	.61	.70	.84		.67	.32	.39
Lower Thoracic	.31	.21	.43	.37	.53	.60	.77	.67		.43	.48
Lateral Arm	.26	.27	.39	.39	.43	.34	.42	.32	.43		.50
Medial Arm	.43	.31	.46	.49	.62	.51	.40	.39	.48	.50	
Mean Correlation	.35	.27	.45	.44	.54	.54	.49	.49	.50	.38	.46
Weight	.36	.19	.33	.33	.47	.43	.61	.61	.54	.36	.43
Fat Weight	.26	.14	.33	.31	.42	.41	.50	.46	.51	.36	.43

Table 4

Correlations of Fat Thickness with Body Weight and Fat Weight. Data Represent Mean Correlations by Body Site for Each of Three Age Groups

	Mean Correlation With:					
	Body Weight			Calculated Fat Weight		
	25-44	45-64	65-84	25-44	45-64	65-84
Forearm	.38	.41	.30	.43	.35	.24
Trunk	.52	.54	.45	.49	.45	.47
Calf	.29	.24	.28	.34	.21	.20
All Sites	.42	.42	.37	.42	.35	.24

Among the 11 radiographic measures of body fat, 8 provided sufficient data and are included in Table 5. Negative increments are observed in measures of arm and calf fat in all the age groups. In the middle trunk, fat lateral to the lowest rib (lower thoracic) shows a negative increment through adulthood. While there were too few greater trochanter level measurements to allow calculation of increments, fat lateral to the anterior-superior iliac spine was available and is about two inches above the greater trochanter sites. Increments at this level were positive in the groups for which sufficient data were available. In the cross-sectional data, the iliac spine site showed intervals of increase and decrease, and its fat pattern was intermediate between that of the trochanter and middle trunk sites.

Correlations were calculated between the fat increments by body site and weight increments, irrespective of age and are shown in Table 6. Of the eight correlations, all but the value for lateral calf fat were significantly positive ($p < .05$). The trunk measures of fat were more highly correlated with weight change than those of either extremity.

DISCUSSION

The decline in weight beginning in early adulthood seen in the cross-sectional data is similar to the finding of constant weight between 25 and 45 years and a decline thereafter in the Normative Aging Study (Damon et al. 1965). Similarly, in the National Health Examination Survey (Stoudt

Table 5

Median Yearly Increments for Body Composition Variables according to Age Decade

	Age Group											
	25-34		35-44		45-54		55-64		65-74		75-84	
	N	Mdn	N	Mdn	N	Mdn	N	Mdn	N	Mdn	N	Mdn
Weight (kg)	10	.303	45	.367	65	.264	52	−.082	43	−.216	14	.226
Height (cm)	10	−.041	45	−.076	65	−.076	52	−.119	43	−.188	14	−.209
Medial Arm Fat	1	—	18	−.015	16	−.029	11	−.029	11	−.077	5	0.0
Lateral Arm Fat	8	−.012	41	−.014	50	−.053	30	−.064	31	−.068	8	−.019
Lower Thoracic Fat	7	.014	19	−.072	32	−.029	29	−.135	28	−.076	6	0.0
Iliac Crest Fat	2	—	3	.523	8	.164	10	−.465	12	.109	2	—
Iliac Spine Fat	2	—	3	.238	17	.525	8	.238	12	.480	2	—
Anterior Calf Fat	9	0.0	42	−.038	54	−.052	37	−.024	32	−.024	8	−.038
Lateral Calf Fat	9	0.0	35	−.054	46	−.045	41	−.026	25	−.018	9	−.115
Medial Calf Fat	2	—	17	0.0	25	−.058	25	−.038	25	−.038	8	−.015
Total Body H_2O %	4	−1.290	16	−.549	36	−.090	26	−.301	22	−.054	7	.961
Total Body Fat %	4	.947	16	.720	36	.123	26	.268	22	−.040	7	−1.310
Extracellular H_2O %	4	−.309	20	−.078	38	−.034	32	−.146	23	.154	7	.310
Intracellular H_2O %	4	−.981	16	−.743	36	−.095	25	−.372	21	−.439	7	−.004
Body Fat Weight (kg)	4	.827	16	.603	36	.184	26	.235	22	−.031	7	−1.140
Fat-free Weight (kg)	4	−.726	16	−.574	36	−.089	26	−.575	22	−.364	7	.644

[1] All fat increments are in millimeters.

Table 6

Correlations between Subcutaneous Fat Increments and Weight Increments

Body Site	N	Correlation with weight increment (r)
Anterior Calf Fat	183	.21
Medial Calf Fat	91	.19
Lateral Calf Fat	166	.07
Iliac Crest Fat	38	.54
Iliac Spine Fat	44	.54
Lower Thoracic Fat	122	.47
Lateral Arm Fat	169	.24
Medial Arm Fat	62	.29

et al. 1965) weight increased by only .5 kilograms between ages 25 and 44 and then began to decline. The finding that body fat percentage increases with age is corroborated by studies of Mickelson (1958), Friis-Hansen (1965), Oleson (1965), and Krzywicki et al. (1967). Although a variety of techniques and calculations have been used in the determination of total body fat, values for adults range between 25 and 35 percent, as in the present findings (Behnke, 1963; Friis-Hansen, 1965; Krzywicki et al. 1967). Intracellular and extracellular water estimates are similar to those of Friis-Hansen (1965). Thus, the indication of rather constant fat weight and declining fat-free weight during adulthood is not unexpected on the basis of previous research. The loss of lean tissue with age has been further substantiated in the Baltimore Longitudinal Study by Tzankoff and Norris (1976), who have shown that creatinine excretion and basal metabolic rate decline with age in this population, indicating loss of muscle mass.

Increases in fat thickness in the lower trunk were expected based on the studies of Brožek (1952), Garn and Young (1956), and Garn (1960). In the present study the greater trochanter site increased in thickness until middle age as Garn and Young had shown. Later in life, however, there is a decline in fat thickness at this site which was not documented by the earlier authors. The decline in fat thickness at the level of the lower thorax was not previously demonstrated and it appears that the increased abdominal girth is not the result of an individual having more adipose tissue.

Thus, fat accumulation occurs primarily on the lower trunk and thighs and, as demonstrated by other studies, on the back as well.

The decline in anterior calf fat corresponds to the findings of Garn and Young (1956). Garn and Saalberg's (1953) finding that there is little change in the medial and lateral fat on the calf was not replicated in the present study, as these sites, along with the posterior calf, decline in early adulthood and then increase later in life. The extent to which this late increase in fat thickness is the result of edema is uncertain, however, subjects in this study are generally in good health and clinical evidence of edema is not common. The lateral and medial aspects of the arm change somewhat differently from one another late in life.

The intercorrelations show the sites on the trunk to be the best indicators of fatness at other body sites. Similar results were reported by Garn (1954) who found that the trochanter sites provide the best intercorrelations and that the medial calf and arm were almost as good. Correlations with weight are best in the middle trunk followed by the lower trunk and arm sites. In keeping with the findings of fat changes occurring primarily in the trunk during later adulthood, the trunk sites are the best correlates in all age groups and the extremity sites become rather poor correlates of body fatness in old age. This shows that in all age groups the abdominal level sites are the best correlates of weight, particularly at the iliac level. Medial arm and leg are good correlates for younger but not older males. Similar findings were reported by Hunt (1961), who used factor analysis to determine which skinfolds are most representative within the body. His results indicated that trunk skinfolds are more representative than limb ones, and that this was especially true in older individuals.

The longitudinal data reveal that sites in close anatomical proximity behave differently during adulthood. Fat in the extremities remains relatively stable in the cross-sectional data, but in the longitudinal data it is observed to decline through adulthood. The middle trunk undergoes a net loss throughout this period and the lower trunk appears to be a site of fat gain. These findings show subcutaneous fat to be in a dynamic state throughout the body during this long age span. The change is most striking in the longitudinal data, because during years of weight and total fat gain in the individual, many subcutaneous fat sites are actually declining. The mechanism by which this change in fat accumulation patterning with age occurs is not revealed by the present or previous studies.

The above findings suggest that, at least during old age, when all subcutaneous fat sites are either declining in thickness or unchanging, total fat remains the same (as shown most clearly in the cross-sectional data). This implies that internal fat is accumulating during a period when subcutaneous fat is decreasing. This was shown by Škerlj et al. (1953, 1959) who calculated the weight of subcutaneous fat from surface area. In a more recent histological study Inokuchi et al. (1975) showed that fat deposits increase in muscle tissue with age. If, as Hirsch (1975) has proposed, the number of adipose cells is established in childhood, subcutaneous fat cells must be decreasing in size in old age while internal fat cells in the abdomen or muscle tissue are increasing in size. This possibility, along with the clear evidence of differential fat accumulation subcutaneously in the body, means that fat cells must be examined with reference to the body site from which they are taken, as all fat cells in the body are not changing in the same way or at the same rate.

ACKNOWLEDGEMENTS

We wish to thank Dr. Stanley Garn of the Center for Human Growth and Development and Department of Anthropology of The University of Michigan for his helpful comments and suggestions during the course of this study. We also express our gratitude to Mr. Robert Wainright for his assistance in data analysis, Dr. A. R. Behnke for consultation and instruction regarding anthropometric measurements, Mr. C. J. Manaras and Mr. J. B. Melvin for measurements of the subjects, and to Mr. M. J. Yiengst and Mrs. E. Beard for supervising the chemical determinations.

LITERATURE CITED

BROŽEK, J. 1952 Change of body composition in man during maturity and their nutritional implication. Fed. Proc. **2:** 784-793.

——— 1961 Body composition. Science, **134:** 920-930.

DAMON, A. 1965 Notes on anthropometric technique. III. Adult weight gain, accuracy of stated weight, and their implications for constitutional anthropology. Am. J. Phys. Anthrop. **23:** 305-311.

DAMON, A., C. C. SELTZER, H. W. STOUDT AND B. BELL 1972 Age and physique in healthy white veterans at Boston. J. Gerontol. **27:** 202-208.

FORBES, G. B. 1976 The adult decline in lean body mass. Human Biol. **48:** 161-173.

FRIIS-HANSEN, B. 1965 Hygrometry of growth and aging. Symp. Soc. Hum. Biol. **7:** 191-209.

GARN, S. M. 1954 Fat patterning and fat intercorrelations in the adult male. Human Biol. **26:** 56-69.

——— 1960 Fat accumulation and aging in males and females. *In:* The biology of aging, B. L. Strehler, (ed.). AIBS Pub. No. 6.

——— 1970 The earlier gain and later loss of cortical bone. Charles C Thomas, Springfield, Illinois.

Garn, S. M. and E. L. Gorman 1956 Comparison of pinch caliper and teleoroentgenogrammetric measurements of subcutaneous fat. Human Biol. **28:** 407-413.

Garn, S. M. and J. H. Saalberg 1953 Sex and age differences in the composition of the adult leg. Human Biol. **25:** 144-153.

Garn, S. M. and Z. Shamir 1958 Methods for research in human growth. Charles C Thomas, Springfield, Illinois.

Garn, S. M. and R. W. Young 1956 Concurrent fat loss and fat gain. Am. J. Phys. Anthrop. **14:** 497-504.

Glanville, E. V. and R. A. Geerdink 1970 Skinfold thickness, body measurements and age changes in Trio and Wajana Indians of Surinam. Am. J. Phys. Anthrop. **32:** 455-461.

Hirsch, J. 1975 Cell number and size as a determinant of subsequent obesity. *In:* Childhood obesity, M. Winick, (ed.). J. Wiley and Sons, New York.

Hooten, E. A. and C. W. Dupertuis 1951 Age changes and selective survival in Irish males. *In:* Studies in physical anthropology, W. W. Howells and S. L. Washburn, (eds.). No. 2 Amer. Assoc. of Phys. Anthrop. and Wenner-Gren Foundation. Pp. 1-129.

Howells, W. W. and H. K. Bleibtreu 1970 Hutterite age differences in body measurements. Peabody Museum Papers **57**(2): 1-123.

Hunt, E. E., Jr. 1961 Measures of adiposity and muscularity in man: some comparisons by factor analysis. *In:* Techniques for measuring body composition, J. Brožek and A. Henschel, (eds.). National Academy of Sciences, National Research Council, Washington, D.C. Pp. 192-211.

Inokuchi, S., H. Ishikawa, S. Iwamoto and T. Kimura 1975 Age-related changes in the histological composition of the rectus abdominis muscle of the adult human. Human Biol. **47:** 231-249.

Krzywicki H. J. and K. S. K. Chinn 1967 Human body density and fat of an adult male population as measured by water displacement. Am. J. Clin. Nutr. **20:** 305-310.

Mickelson, O. 1958 Age changes in body composition. Public Health Reports, **73:** 295-301.

Montoye, H. J., F. H. Epstein and M. O. Kjelsberg 1965 The measurement of total body fatness: a study in a total community. Am. J. Clin. Nutr. **16:** 417-427.

Moore, F. D., K. H. Oleson, J. D. McMurrey, H. V. Parke, M. R. Bell and C. R. Boyden 1963 The body cell mass and its supporting environment. Saunders, Philadelphia.

Norris, A. H., T. Lundy and N. W. Shock 1963 Trends in selected indices of body composition in men between the ages 30 and 80 years. Ann. N.Y. Acad. Sci. **110:** 623-639.

Oleson, K. H. 1965 Body composition in normal adults. *In:* Human body composition: approaches and applications, J. Brožek (ed.). Pergamon Press, Oxford. Pp. 177-190.

Škerlj, B. 1959 Age changes in fat distribution in the female body. Acta Anat. **38:** 56-63.

Škerlj, B., J. Brožek and E. E. Hunt, Jr. 1953 Subcutaneous fat and age changes in body build and body form in women. Am. J. Phys. Anthrop. **11:** 577-600.

STONE, J. L. AND A. H. NORRIS 1966 Activities and attitudes of participants in the Baltimore Longitudinal Study. J. Geront. **21**: 575-580.

STOUDT, H. W., A. DAMON, R. A. MCFARLAND AND J. ROBERTS 1965 Weight, height and selected body measurements of adults. United States, 1960-62. U.S. Public Health Service Publication No. 1000, Series 11, No. 8. Government Printing Office, Washington, D.C.

——— 1970 Skinfolds, body girths, biacromial breadth and selected anthropometric indices of adults. United States, 1960-62. U.S. Public Health Service Publication No. 1000, Series 11, No. 35. Government Printing Office, Washington, D.C.

TZANKOFF, S. P. AND A. H. NORRIS 1976 The constancy of the basal metabolic rate with age. Abstract. *In:* Program of the 29th Annual Meeting of the Gerontological Society, October 13-17, New York, p. 36.

Experimental Aging Research, Volume 8, Number 1, 1982

Neuroticism, Coronary Artery Disease, and Chest Pain Complaints: Cross-Sectional and Longitudinal Studies[1]

PAUL T. COSTA, JR.[2], JEROME L. FLEG[2],
ROBERT R. MCCRAE[2] AND EDWARD G. LAKATTA[2]

Cardiovascular disease increases with age, but some of the factors thought to be related to CAD, including the personality disposition of neuroticism, show a pattern of lifelong stability. In the present paper, chest pain reports and other CAD symptoms are analyzed in the context of a model of personality and health which emphasizes the role of the enduring disposition of neuroticism as a determinant of perceptions of health. In the first study, the association between chest pain complaints and psychological distress measured by depression and low general well-being is documented for both sexes and for age groups from 25 to 74 in a national sample. Chest pain complaints increased with age, but an hypothesized age by chest pain group interaction was not found. In the second study, the direction of the causal connection is investigated in a longitudinal retrospective-predictive study of CAD diagnoses in a sample of 123 men followed for periods of up to 20 years. Personality variables did not predict the development of CAD, but emotional instability and diffuse somatic complaints were predictive of anginal diagnoses in subjects who showed no other signs of CAD. Alternate interpretations of the differential prediction of anginal complaints from neuroticism are discussed.

An extensive literature spanning the fields of medical sociology, epidemiology, medical psychology, and behavioral medicine has consistently shown that self-perceptions of health are related to various psychological characteristics such as health attitudes, morale, adjustment, and psychological distress. There have been incisive descriptions of the role of social psychological factors, especially psychological distress, in affecting the presentation of bodily complaints, perceptions of personal vulnerability, and utilization of medical care [18; 19; 23]. While much evidence shows that the individual's psychological state influences his or her sense of physical well-being, recognition of the central role of the normal personality dimension of neuroticism has been slow to emerge.

Conceptually, many of the psycho-social variables related to health perceptions, including distress, personal vulnerability, and bodily concerns can be seen as manifestations of neuroticism, a dimension of personality usually measured by anxiety, depression, or hostility scales. Empirical studies have shown that variables such as low well-being and hopelessness [5] and poor attitudes toward health [7] can be predicted from measures of neuroticism over periods of up to ten years. At the same time, neuroticism itself has been shown to predict a wide spectrum of medical complaints and symptoms [6]. On the other hand, large-scale studies [16] have provided compelling evidence that neuroticism is not related to excess or premature mortality.

The present paper addresses the two questions of whether neuroticism is a predictor of anginal diagnoses and whether it is a risk factor for CAD. In posing these two issues as separate questions, we emphasize that it is important to distinguish between the two major clinical manifestations of coronary artery disease: myocardial infarction (MI) and angina pectoris (AP). Infarction signifies destruction of some portion of the ventricular muscle, due to interruption of its blood supply, whereas angina pectoris usually indicates reversible myocardial ischemia. While it is true that an MI normally presents with dramatic clinical symptoms, a sizeable percentage of infarctions are asymptomatic or "silent"; thus physicians most often rely on electrocardiographic evidence (QRS complex change), serum enzyme patterns and other diagnostic studies of the heart to diagnose MI. Angina pectoris, on the other hand, is a symptom, i.e., a report of recurrent characteristic chest pain usually elicited by exercise or emotional arousal and relieved by rest or nitrates. Because of its characteristic symptomatology, the diagnosis is most commonly made by clinical history given by the patient. Often, there is no objective physical or ECG finding at rest in patients with angina but no history of MI.

Jenkins, in comprehensive reviews in 1971 and again in 1976, found 45 published reports dealing with psychological precursors of CAD [13; 14]; however, most of these studies disregarded the distinction between angina and other manifestations of CAD. In a more recent review of behavioral risk factors in CAD [15], Jenkins wrote that neuroticism was associated with CAD, but he noted that the variables of anxiety and neuroticism were more consistently related to angina pectoris than to myocardial infaction. Jenkins concluded that, along with life dissatisfaction, family problems, and "psychosocial discord," " the anxiety-neuroticism dimension represents one more risk factor that operated differently for angina than for infarction" (p. 554).

The perspective offered here suggests an alternative interpretation of the data. Instead of seeing neuroticism as a risk factor for CAD, it might better be interpreted as a determinant of health perceptions and symptom reports, which lead in a certain number of cases to diagnoses of AP. It can be hypothesized that individuals high in neuroticism will be more likely to be diagnosed as having AP at some time in their life, but will be no more likely than others of their age to show objective evidence of CAD.

---

[1] Requests for reprints should be sent to Paul T. Costa, Jr., Chief, Section on Stress and Coping, Gerontology Research Center, Baltimore City Hospitals, Baltimore, MD 21224. Some of the data in this paper were provided by the National Center for Health Statistics; however, analyses, interpretations and conclusions based on these data are the sole responsibility of the authors, not the NCHS. The assistance provided by Dr. Hal J. Dupuy in making these data available is gratefully acknowledged.

[2] From the Gerontology Research Center, National Institute on Aging, NIH, Baltimore City Hospitals, Baltimore, MD 21224, U.S.A.

The question which this paper addresses is whether, in normal unscreened populations, the personality dimension of neuroticism influences the individual's reporting of symptoms, especially chest pain, thereby influencing the diagnosis of CAD. The question is approached in two parts. The first examines whether self-reported chest pain descriptions are related to a measure of psychological distress and to depression in a large national probability sample of men and women. The second examines the retrospective and predictive relations between CAD symptom presentation and personality traits in a group of male participants in the Baltimore Longitudinal Study of Aging.

Study 1

The association of medical complaints with psychological distress is a widely recognized phenomenon known through clinical experience and a few large scale studies. There is, however, no direct evidence that specific questions about chest pains follow this same pattern. The first study examines the relation between self-reported chest pains and two measures of psychological distress in a national probability sample of over 6000 men and women. It is hypothesized that individuals who report various kinds of chest pain will be lower in general well-being and higher in depression than are individuals who report no chest pain.

As clinical cardiologists know, chest pain does not always imply heart disease. This is particularly likely to be true among older age groups, where decreased activity levels may mask angina, and non-coronary diseases may mimic angina [11]. If psychological distress is specifically linked to CAD, then both these effects will tend to attenuate the observed relation. For this reason, it can be hypothesized that there will be an age by chest pain group interaction, such that the association of chest pain with psychological distress will be higher among younger subjects.

Method

Subjects. Data for this study were collected as part of the Health and Nutrition Examination Survey (HANES) conducted by the National Center for Health Statistics during the period from 1971 to 1975. Subjects were part of a national probability sample of non-institutionalized men and women aged 25-74, and data were collected from 100 locations in the coterminous United States. The General Well-Being scale (GWB) was administered to 6913 subjects from the 100 locations; the Center for Epidemiological Studies depression scale (CES-D) was administered to 2814 subjects from the last 36 locations surveyed [8]. Details on the sampling procedures used are available from NCHS. In general, however, the sample may be regarded as representative of the non-institutionalized U.S. population for the given age range.

Measures. The GWB is an 18-item measure of subjective well-being which contains subscales on freedom from health worry, energy level, satisfying or interesting life, cheerful vs. depressed mood, relaxed vs. tense or anxious, and emotional behavior control. It has been interpreted as a measure of nonspecific distress [17], and correlates in a college population with measures of depression and anxiety. It might therefore be regarded as a measure of neuroticism. Internal consistency (coefficient alpha) is .91 in this sample.

The CES Depression scale is a 20-item self-report survey which covers psychic expressions of depression including hopelessness, sadness, loneliness, and anxiety. Poor appetite, crying spells, and sleep disturbance are also measured. The GWB and CES Depression scales show a correlation of $r = -.72$ in the present sample.

Three self-report items on chest pain were asked of subjects: Have you ever had: Trouble with any pain or discomfort in your chest? Trouble with any pressure or heavy sensation in your chest? Severe pain across the front of your chest lasting for half an hour or more? Responses were recorded as "yes" or "no": 17.3% of subjects reported discomfort; 13.7% reported pressure; and 7.6% reported severe pain.

Analyses. Three-way analyses of variance were conducted using 10 levels of age (five year intervals), sex, and response to the chest pain questions as the classifying variables; and GWB and Depression scores as the dependent variables. Consistent with previous findings [6], the proportion of individuals reporting chest pain was higher in the older groups. Because of the large sample size and the unequal cell frequency, a stringent probability level of .001 was required for significance.

Results

Subjects who reported pain or discomfort in the chest were lower in General Well-Being ($F = 553.85$) and higher in CES-Depression ($F = 90.24$) than those with no pain. Subjects with pressure or a heavy sensation in the chest were also lower in General Well-Being ($F = 459.67$) and higher in depression ($F = 86.92$). Finally, the smaller group of subjects who reported severe pain across the front of the chest, lasting for half an hour or more, showed an identical pattern, being lower on GWB ($F = 239.59$) and higher in depression ($F = 56.81$). The magnitude of these differences is indicated in Figure 1, which shows the GWB scores for the subjects with and without pain or discomfort in chest at each age group. The magnitude of differences is notable, amounting to about two-thirds of a standard deviation on the GWB scale, and about one-half of a standard deviation on the depression scale.

In all analyses, sex and symptom reporting were significant ($p < .001$) main effects, whereas age was never significant. Women showed significantly lower General Well-Being and significantly higher depression scores, although the magnitude of the effect for sex differences was not as large as the effect for chest pain classification. None of the two or three way interactions reached the .001 level of significance. Thus, there was no support for the hypothesis that age mediated the relationship between distress and chest pain.

Since the GWB scale contains a subscale concerned with health, it is possible that realistic health concern expressed in this single subscale is responsible for all the observed differences. In order to test this hypothesis, additional analyses were conducted for each of the six subscales of the GWB. Results showed that subjects who reported any of the three kinds of chest pain were significantly lower in all six of the GWB subscales, not merely the health concern subscale.

These analyses were conducted on a representative sample of the noninstitutionalized population of the United States, aged 25-74. However, since this country is predominantly white, it is possible that the overall results are characteristic only of whites. The large sample size allowed a replication of analyses restricted to non-whites. In these analyses, age of the subjects was not used as a classifying factor, since in the first analysis age was significant neither as a main effect nor in interaction with other classifying variables. Results from these analyses exactly paralleled the full sample findings: non-white

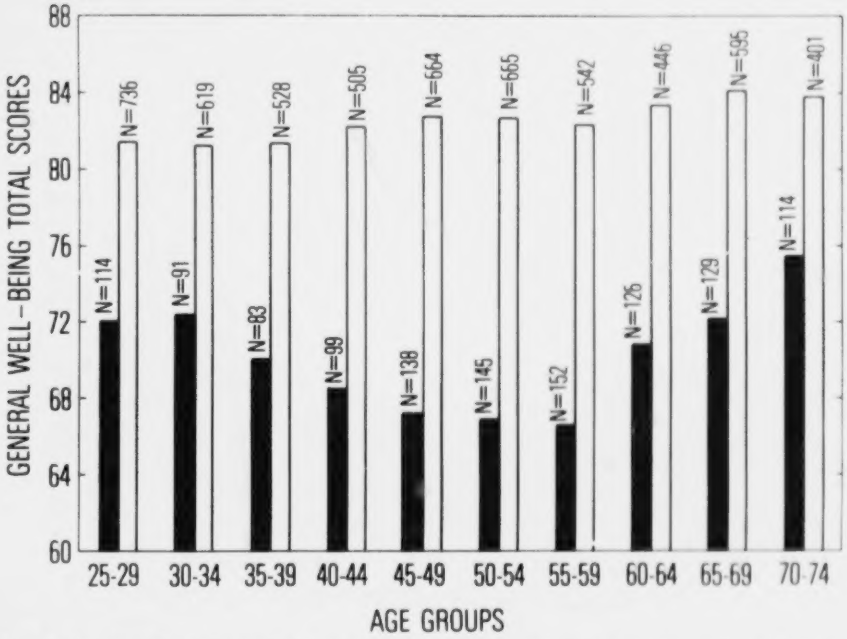


Figure 1. Mean General Well-Being scores for subjects who did (solid) or did not (open) report "pain or discomfort in chest" at each age

subjects who reported chest discomfort, pressure, or severe chest pain were lower in total GWB and on each of the GWB subscales, and higher in Depression. The same sex differences were also found, and again there were no interactions.

Study 2

Although the previous results demonstrate rather conclusively the association of perceived chest pain with measures of psychological distress (or neuroticism) in a large and representative sample, the interpretation of the finding is obscured by a number of factors. Of these, the causal ordering of the relation is most ambiguous. It could plausibly be argued that depression and lowered well-being are the results of CAD, or the attendant angina, or the knowledge of the individual that he or she probably has a life-threatening illness.

Another problem in interpreting these data is that chest pain complaints are not equivalent to angina. The examining physician reaches a judgment of probable or definite angina only after detailed questioning of subjects on the origin, phenomenology, and duration of chest pain, and the methods which the subject has found useful in reducing it. Many reports of chest pain are properly discounted by the physician as being non-anginal or non-ischemic. Physician's expertise is particularly necessary for older individuals, in whom health problems other than CAD complicate diagnosis.

In order to examine these alternative interpretations of the data, medical records of male participants in the Baltimore Longitudinal Study of Aging were examined. Personality and self-reported health data were collected prior to the presentation of any evidence of CAD. Individuals who subsequently developed angina and/or electrocardiographic (ECG) signs of CAD were compared with others who remained free of CAD over an extensive follow-up period. This study thus combines physician's evaluations of angina with longitudinal data which allow a much stronger basis for the inference of causal relations between angina and neuroticism.

This second study attempts to answer two specific questions: (1) is neuroticism a risk-factor for CAD – i.e., are there pre-existing differences in emotional stability and health perceptions between individuals who will and those who will not subsequently develop CAD; and (2) do pre-existing differences in neuroticism and or somatic complaints predict anginal symptom reporting among individuals who will be diagnosed as having CAD?

Method

Subjects. Subjects in this study were volunteers in the Baltimore Longitudinal Study of Aging who have received medical examinations and psychological tests during the period from 1958 to the present. The 123 subjects used in this study ranged in age from 33 to 84 at entry into the study. Participants in the BLSA are a well-educated, community-dwelling group who are somewhat higher than average on measures of emotional-stability. Sample characteristics are discussed further elsewhere [7].

Procedures. Subjects receive intensive medical examination at the BLSA on each visit, every one to two years. In addition, resting ECG's are obtained routinely. Double Masters (1958-1968) and treadmill exercise testing (1968 to present) were also given to subjects not classified as having definite CAD.

On the basis of these medical records, subjects were classified into several groups:

1. The *ECG Signs without Anginal Symptoms (Asymptomatic)* group consisted of 27 men, free of ECG signs, anginal symptoms, or history of MI at entry, who during the course of the study, showed definite signs of CAD, using resting or stressed ECG ST segment depressions as criteria. However, these men reported no history of MI and had no angina in a follow-up period of from 5 to 15 years. They remained free of these symptoms at their last visit to the BLSA.

2. The *Anginal Symptoms without Signs or History (Anginal-Symptoms-Only or ASO)* group consisted of 10 men, free from ECG signs, anginal symptoms, or history of MI at entry. During the course of the study, they were diagnosed as having probable or definite angina. However, these men never had an MI nor ECG signs of ischemia in a follow-up period of from 5 to 11 years, and remained free of these signs at their last visit to the study.

3. The *CAD with both ECG Signs and Anginal Symptoms (CAD-with-Both)* group consisted of 49 men free of ECG signs, anginal symptoms, or history of MI at entry into the study. In the course of the study, they were diagnosed as having probable or definite angina, and in addition, they showed either definite ECG signs of MI or IHD, or a history of MI.

4. The *Reference or Control (Control)* group consisted of 29 men, selected from several hundred, free of ECG signs, anginal symptoms, and history of MI at entry, and at last visit to the BLSA, 10 to 20 years later. Although free of

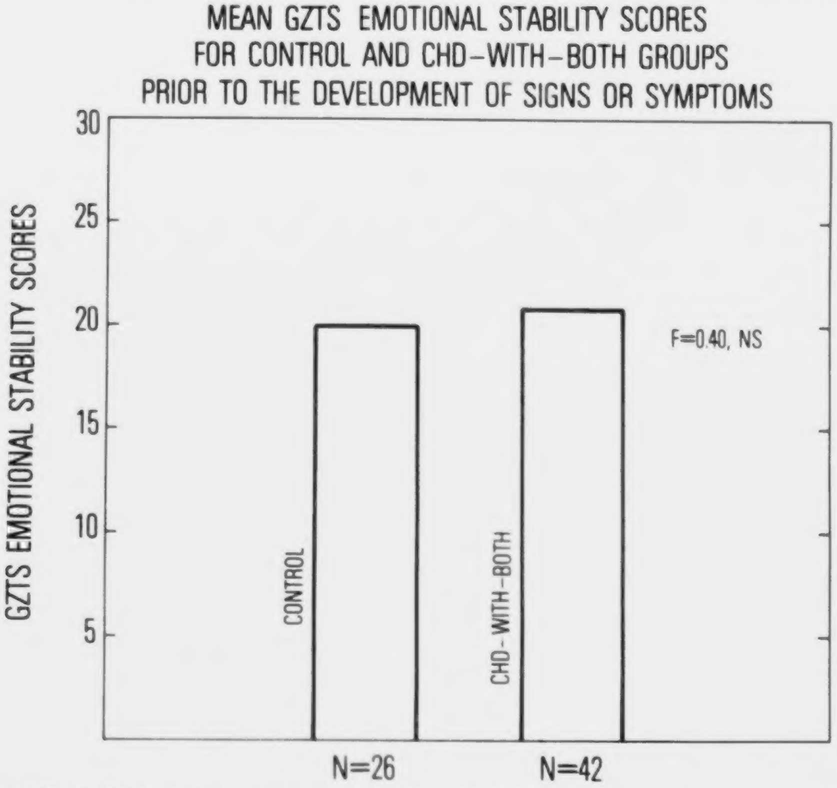


Figure 2. Mean GZTS Emotional Stability scores for subjects who subsequently developed both CAD signs and symptoms versus controls who developed neither.

> CAD, no attempt was made to screen these subjects for other illnesses. These subjects were roughly age matched to the other three groups, but were otherwise randomly selected from the population of subjects free from CAD, and with sufficient follow-up interval.

All subjects were given the Guilford-Zimmerman Temperament Survey (GZTS) [12] and the Cornell Medical Index (CMI) [3] at their first or second visit. In every case, administration of these tests preceded the first appearance of ECG signs or anginal symptoms by at least one year.

Analyses. Since increasing age is a risk factor for CAD, some control for its influence was necessary. The small number of subjects in this study precluded analyses within age groups, so age was used as a covariate in all analyses reported.

In order to test the hypothesis that emotional instability predisposes the individual to the development of CAD, analysis of variance (with age as a covariate) was conducted contrasting the clearest CAD group, the CAD-with-both subjects, with the Control subjects. It was specifically hypothesized that Control subjects would be higher on the GZTS Emotional Stability scale. In addition, differences in the nine other GZTS scales were also examined.

In order to test the hypothesis that personality differences would predict differences in CAD signs and symptoms, analysis of variance was conducted using CAD group (Asymptomatic, Anginal-Symptoms-Only, and CAD-with-Both) as the classifying factor. It was hypothesized that Asymptomatics would be highest and Anginal-Symptoms-Only subjects lowest in emotional stability. It was also hypothesized that a tendency toward bodily concern, as measured by CMI sections and total physical complaints, would differentiate the groups. Specifically, Anginal-Symptoms-Only subjects, who report chest pain in the absence of ECG signs of ischemia, were hypothesized to be highest in somatic complaints in the period before the development of angina.

Finally, in order to replicate results with a larger number of Angina-Only subjects, analyses were repeated including subjects who reported angina on entry into the study among the Anginal-Symptoms-Only group. Note that this analysis, unlike all the others, is not strictly a retrospective-predictive study.

Results

Comparison of individuals who subsequently developed CAD with angina and individuals who remained free of both CAD signs and angina over the follow-up period of from 10 to 20 years showed no significant differences in emotional stability ($F = 0.40$, *n.s.*). Figure 2 shows that both groups began with comparable levels of distress-proneness. *F*-tests also indicated that there were no pre-existing differences in any of the other GZTS scales, measuring such traits as sociability, masculinity, activity, ascendency, and friendliness-hostility. These data provide no support for the hypothesis that temperaments measured by the GZTS play a causal or etiological role in the development of CAD. In addition, total physical complaints as measured by the CMI also fail to differentiate CAD-with-Both subjects from Control subjects ($F = 0.05$, *n.s.*). It thus seems unlikely that CMI sums represent early signals of the development of CAD.

When comparisons are made among CAD groups, however, a definite pattern emerges. As Figure 3 shows, Anginal-

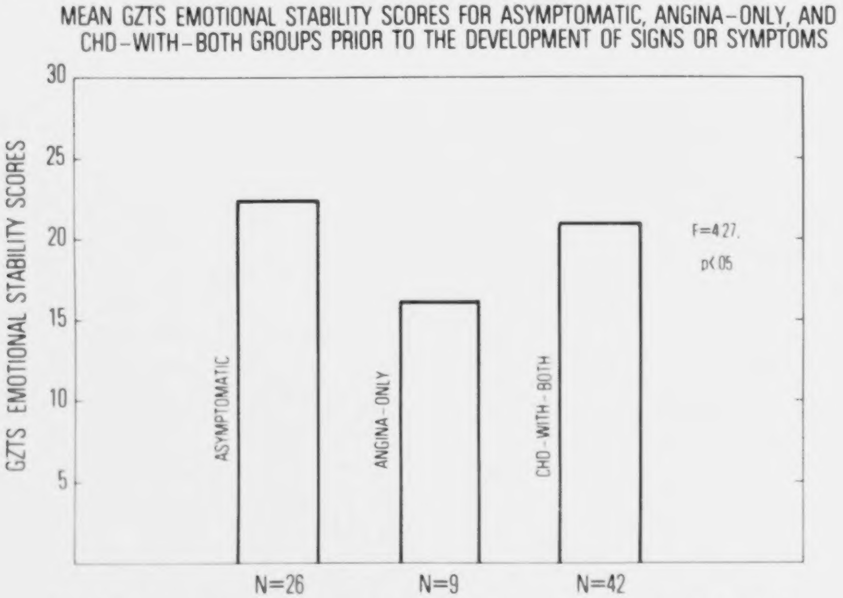


Figure 3. Mean GZTS Emotional Stability scores for three groups of subjects subsequently diagnosed as having CAD

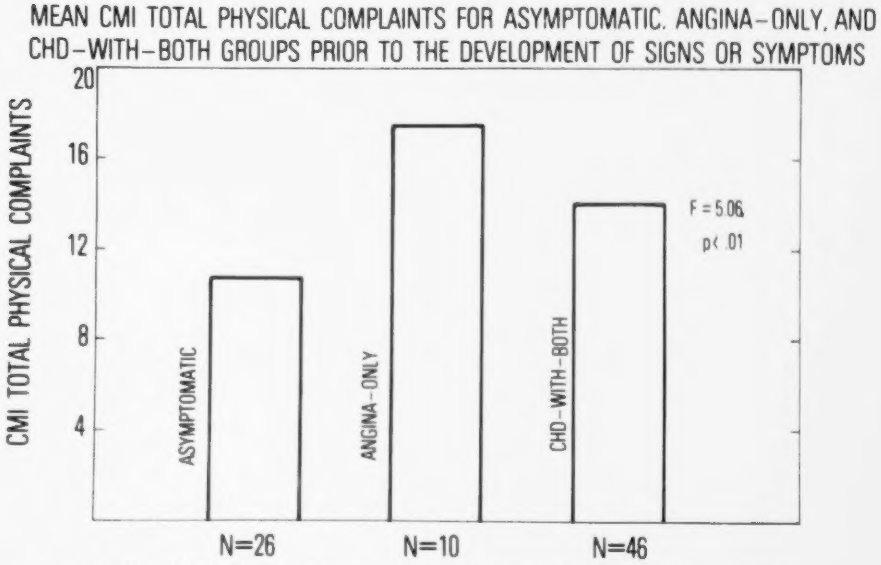


Figure 4. Mean CMI Total Physical Complaints for three groups of subjects subsequently diagnosed as having CAD.

Symptoms-Only subjects are lowest in Emotional Stability and Asymptomatic subjects highest in Emotional Stability ($F = 4.27$, $p < .05$). Anginal-Symptoms-Only subjects are also highest, and Asymptomatic subjects lowest, in total number of reported physical complaints on the CMI ($F = 5.06$, $p < .01$), as shown in Figure 4. Post-hoc comparisons reveal that the Asymptomatic subjects did not differ significantly from the CAD-with-both groups on these measures.

In addition, analyses on the nine other GZTS scales, and on the specific sections of the CMI show that Asymptomatics are significantly more restrained in personality, and less likely to complain of cardiovascular or digestive problems. They also have better health habits. In each of these cases, the Anginal-Symptoms-Only group had the most cardiovascular and digestive complaints, and the poorest health habits.

Only ten subjects were initially free of CAD symptoms and subsequently developed angina without other evidence of CAD. As a result, statistical comparisons are not as powerful as might be hoped. In order to check the robustness of the observed effects, the same analyses were performed using a different criterion for membership in the Anginal-Symptoms-Only group. In addition to those individuals initially free of CAD signs and symptoms, eight individuals who entered the study with anginal complaints, but showed neither ECG signs nor history of MI during a follow-up period of from 5 to 11 years were also included. These analyses are therefore not strictly retrospective predictive, since we cannot be sure that personality antedated anginal complaints for these eight men. If we assume, on the basis of previous results, that any personality differences are probably pre-existing, then these supplementary analyses are useful as statistically stronger tests of the same hypotheses.

As might be expected, Anginal-Symptoms-Only subjects were significantly higher in cardiovascular complaints on the CMI – a finding which might be attributed to the angina of some of the members of the group. However, these same subjects were also significantly higher on complaints in the digestive, respiratory, musculo-skeletal, and skin sections, as well as the total physical complaints ($F = 6.44$, $p < .01$).

Once again, the Asymptomatic subjects were highest and the Anginal-Symptoms-Only subjects lowest, in emotional stability ($F = 6.18$, $p < .01$). However, in these analyses, the effect for restraint was non-significant.

Discussion

Several other studies confirm the present pattern of results. An early, carefully executed – but largely ignored – study was that of Ostfeld, et al. [21]. They prospectively studied 1885 men over a four and one-half year period, using the 16PF and the MMPI to investigate personality factors contributing to CAD. They contrasted the personality scores of men who had and those who had not developed CAD and between those who developed only AP and those men who had only MI. They reported that the Hs, K, and Hy scales from the MMPI along with the Emotional Stability scale from the 16PF (all measures of neuroticism) were elevated in men before the development of AP; whereas men who were to develop MI were not different from those without coronary disease on any of the scales.

Further confirmation comes from a prospective study of AP among 10,000 Israeli men by Medalie, et al. [17]. The incidence of angina was twice as great among men scoring in the top half of the sample on a three item anxiety index. The anxiety score was not predictive of MI, however.

The association of psychological distress or neuroticism with self-reports of chest pain and physicians' diagnoses of angina pectoris is interpretable in several different ways. Four possibilities deserve careful consideration: (1) neuroticism may be a risk factor for the development of CAD; (2) neuroticism may sensitize the individual to pain, and so be part of an "early warning system" for CAD; (3) neuroticism may be related to some other disease which mimics CAD, such as neurocirculatory asthenia (NCA); or (4) neuroticism may lead to medically groundless, though psychologically significant, hypochondriacal complaints.

It is tempting to consider neuroticism as a risk factor for CAD, and it has been so classified by some writers [15]. After all, neuroticism predicts the development of angina, and anginal symptoms are known to be the best non-invasive evidence of CAD available to the physician. The belief that anxiety, depression, and hostility and their psychophysiological manifestations can induce organic illness in individuals is almost universally subscribed to in some form, and is a fundamental concept in psychosomatic medicine. Thus, the "risk factor" interpretation is appealing.

But there are compelling reasons to doubt this interpretation of the data. Since neuroticism predicts angina and angina predicts MI, it follows that neuroticism should predict MI. Yet the Ostfeld and Medalie studies provide strong evidence that this is not the case. In the present study, the individuals in the CAD-With-Angina group, who most clearly suffered from heart disease, did not differ from disease-free controls in antecedent levels of neuroticism. Finally, in a definitive study of 9000 psychoneurotics and 9000 controls followed up after twenty-four years, Keehn, Goldberg, and Beebe [16] reported that there were no differences in CAD-related mortality between the two groups.

These findings strongly suggest that some portion of the diagnoses of AP are false positives, and that it is the false positive group which accounts for the elevated neuroticism among angina patients. More direct evidence is obtainable from recent studies which have looked at psychological characteristics of individuals undergoing coronary arteriography. If neuroticism led to the development of ischemic heart disease, then those individuals who are highest in neuroticism should show the greatest degree of stenosis, or occlusion of the arteries. However, recent studies [2; 25] have failed to find a significant association between the two.

Even more striking are the findings of another study by Elias, et al. [9]. A group of 136 men and women scheduled for arteriography were asked to complete several psychological measures of anxiety, depression, and somatic complaints while awaiting the procedure in the hospital. Understandably, considering the immediate circumstances, the group as a whole had higher levels of anxiety and depression than reference groups. However, when degree of maximum stenosis was correlated with the psychological measures a significant *negative* correlation was found. The more anxious, depressed, or concerned with somatic complaints the individual was, the healthier his or her coronary arteries.

This finding, of course, does not mean that worrying is good for one's health. Instead, it can be understood in terms of the selection processes involved in gathering a sample of individuals about to undergo arteriography. Individuals scheduled for an expensive and risky procedure like arteriography are either medically ill or severe and persuasive complainers. Since neuroticism does not appear to be related to CAD, the medically ill group will contain individuals from the full range of neuroticism. However, among the false positives there will be a disproportionate number of persons high in neuroticism, since neuroticism is known to be associated with somatic complaints [6]. Neuroticism, it appears, is a "risk factor" for undergoing arteriography, but not for myocardial ischemia.

It is possible, however, that neuroticism may be linked to the early detection of CAD. If individuals high in neuroticism have a lower threshold for the detection of pain, or if they attend more closely to their own physiological condition than others, then they may notice the signs of CAD earlier. We do not currently know whether pain sensitivity is higher in such people, but we do know that they are more willing to report pain [6], and are more likely to visit their physician [18; 24]. These behaviors could lead to earlier diagnosis of CAD in individuals who have developed the disease from whatever causes. As Steptoe [22] notes, "The association [of neuroticism] with angina may be non-specific; neurotic individuals are more likely to report symptoms of any origin, and their heart disease will consequently be identified at an earlier stage" (p. 165).

Somewhat similar findings are in fact reported in a Swedish study on diagnosed and undiagnosed hypertension [1]. Hypertension is usually a silent disease, discovered only on physical examination. Yet data show that among Swedish men, those with traits akin to neuroticism are more likely to have been diagnosed and treated. Very well-adjusted individuals often ignore or minimize minor pains, and may neglect medical attention when they should seek it. In this respect, neuroticism may be of benefit to those who can profit from timely diagnosis. Additional studies are required on this question, as well as the related question of how well individuals with different personality characteristics adhere to medical regimens once their illnesses are discovered.

It has long been suggested that many individuals who complain of heart troubles may be suffering from a non-lethal disease called neurocirculatory asthenia (NCA), in which disturbing physiological responses of the heart are triggered by psychological states. This condition is found more frequently in women than in men [4], as is neuroticism [10]. It may be the case that NCA-related angina is predicted by neuroticism, whereas CAD-related angina is not. Fortunately, classical NCA is distinguishable from CAD by a pattern of symptoms, as well as by negative findings on coronary angiography. Research now in progress on patients awaiting angiography involves collecting detailed information on personality and symptom presentation; thus we may be able to resolve this issue.

Finally, it may be that for some individuals chest pain, like complaints about other organs, is simply a sign of emotional distress, unrelated to any organic condition. If so, then epidemiological studies should take this fact into account in estimating prevalence of the disease. The HANES data on self-reported chest pain, for example, probably markedly overestimates the real prevalence of the disease, especially in women. Theoretically, too, this distinction is of great importance, since it represents a direct challenge to certain psychosomatic hypotheses, and obviates research on a link between emotional states and this illness. However, on the level of the individual patient, chest pain is a sufficiently important indicator of CAD that the physician cannot afford to ignore it, even in an individual obviously high in neuroticism. Maladjusted individuals are as susceptible to CAD as anyone else, and their chest pain complaints should not be dismissed as non-ischemic solely on the basis of their personality traits. In cases where detailed medical evaluation shows that the chest pain is not a symptom of an organic condition, the physician should attempt to relieve the psychological distress of the patient.

References

1. Berglund, G., Ander, S., Lindström, B., & Tibblin, G. Personality and repor

ting of symptoms in normo- and hypertensive 50-year-old males. *Journal of Psychosomatic Research*, 1975, *19*, 139-145.

2. Blumenthal, J.A., Thompson, L.W., Williams, R.B., & Kong, Y. Anxiety proneness and coronary heart disease. *Journal of Psychosomatic Research*, 1979, *23*, 17-21.

3. Brodman, K., Erdmann, A.J., & Wolff, H.G. *The Cornell Medical Index health questionnaire manual*. New York: Cornell University Press, 1960.

4. Caranasos, G.J. Neurocirculatory asthenia. In R.S. Eliot (Ed.), *Stress and the heart*. New York: Futura, 1974.

5. Costa, P.T., Jr., & McCrae, R.R. The influence of extraversion and neuroticism on the subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, 1980, *38*, 668-678.

6. Costa, P.T., Jr., & McCrae, R.R. Somatic complaints in males as a function of age and neuroticism: A longitudinal analysis. *Journal of Behavioral Medicine*, 1980, *3*, 245-257.

7. Costa, P.T., Jr., McCrae, R.R., & Norris, A.H. Personal adjustment to aging: Longitudinal prediction from neuroticism and extraversion. *Journal of Gerontology*, 1981, *36*, 78-85.

8. Dupuy, H.J. The psychological section of the current Health and Nutrition Examination Survey. In *Proceedings of the public health conference on records and statistics meeting jointly with the national conference on health statistics*. Washington, DC: National Center for Health Statistics, 1972.

9. Elias, M.F., Robbins, M.A., Blow, F.C., Rice, A.P., & Edgecomb, J.L. Symptom reporting, anxiety and depression in arteriographically classified middle-aged chest pain patients. *Experimental Aging Research*, 1982, *8*, 45-51.

10. Eysenck, S.B.G., & Eysenck, H.J. Scores on three personality variables as a function of age, sex and social class. *British Journal of Social and Clinical Psychology*, 1969, *8*, 69-76.

11. Fleg, J.L., & Lakatta, E.G. Coronary artery disease in the elderly. *Behavioral Medicine Update*, 1981, *3*, 11-14.

12. Guildford, J.S., Zimmerman, W.S., & Guilford, J.P. *The Guilford-Zimmerman Temperament Survey Handbook: Twenty-five years of research and application*. San Diego, CA: EDITS Publishers, 1976.

13. Jenkins, C.D. Psychologic and social precursors of coronary disease. *New England Journal of Medicine*, 1971, *284*, 244-255, 307-317.

14. Jenkins, C.D. Recent evidence supporting psychologic and social risk factors for coronary disease. *New England Journal of Medicine*, 1976, 987-994, 1033-1038.

15. Jenkins, C.D. Behavioral risk factors in coronary artery disease. *Annual Review of Medicine*, 1978, *29*, 543-56.

16. Keehn, R.J., Goldberg, I.D., & Beebe, G.W. Twenty-four year mortality follow-up of army veterans with disability separations for psychoneurosis in 1944. *Psychosomatic Medicine*, 1974, *36*, 27-46.

17. Link, B., & Dohrenwend, B.P. Formulation of hypotheses about the true prevalence of demoralization in the United States. In B.P. Dohrenwend, B.S. Dohrenwend, M.S. Gould, B. Link, & R. Wunsch-Hitzig (Eds.), *Mental illness in the United States: Epidemiologic estimates*. New York: Praeger, 1980.

18. Mechanic, D. Sex, illness, illness behavior, and the use of health services. *Journal of Human Stress*, 1976, *2*, 29-40.

19. Mechanic, D. Stress, illness, and illness behavior. *Journal of Human Stress*, 1976, *2*, 2-6.

20. Medalie, J.H., Snyder, M., Groen, J.J., Neufeld, H.N., Goldbourt, U., & Riss, E. Angina pectoris among 10,000 men: 5 year incidence and univariate analysis. *American Journal of Medicine*, 1973, *55*, 583-594.

21. Ostfeld, A.M., Lebovits, B.Z., Shekelle, R.B., & Paul, O. A prospective study of the relationship between personality and coronary heart disease. *Journal of Chronic Diseases*, 1964, *17*, 265-276.

22. Steptoe, A. *Psychological factors in cardiovascular disorders*. New York: Academic Press, 1981.

23. Tessler, R., & Mechanic, D. Psychological distress and perceived health status. *Journal of Health and Social Behavior*, 1978, *19*, 254-262.

24. Tessler, R., Mechanic, D., & Dimond, M. The effect of psychological distress on physician utilization: A prospective study. *Journal of Health and Social Behavior*, 1976, *17*, 353-364.

25. Zyzanski, S.J., Jenkins, C.D., Ryan, T.J., Flessas, A., & Everist, M. Psychologic correlates of coronary angiographic findings. *Archives of Internal Medicine*, 1976, *136*, 1234-1237.

Journal of Behavioral Medicine, Vol. 3, No. 3, 1980

Somatic Complaints in Males as a Function of Age and Neuroticism: A Longitudinal Analysis[1]

Paul T. Costa, Jr.,[2,3] and Robert R. McCrae[2]

Accepted for publication: March 10, 1980

Previous research has shown that both age and neuroticism are correlated with total scores on self-report health inventories; the present study concerns the influence of these two factors on reports of physical complaints in various bodily systems. Six- and twelve-year longitudinal analyses of the physical health sections (A–L) of the Cornell Medical Index were supplemented with cross- and time-sequential analyses. Subjects, aged 17–97, were taken from a group of 1038 male participants in the Baltimore Longitudinal Study of Aging. Results showed that problems in sensory, cardiovascular, and genitourinary systems increased with age, while health habits improved. More neurotic subjects, as measured by the psychiatric sections (M–R) of the CMI and the Emotional Stability Scale of the GZTS showed higher levels of endorsements on all sections. These results suggest that age does not produce a generalized increase in physical complaints; instead, specific age-related symptoms show increases. Implications of these findings for research involving self-assessments of health are discussed.

KEY WORDS: age; neuroticism; somatic complaints; longitudinal analyses; psychological distress.

---

[1] Portions of this paper were presented at the 32nd Annual Meeting of the Gerontological Society, November 1979, Washington, D.C.

[2] Gerontology Research Center, National Institute on Aging, National Institutes of Health, Baltimore City Hospitals, Baltimore, Maryland 21224.

[3] To whom correspondence should be addressed at Section on Stress and Coping, Gerontology Research Center, National Institute on Aging, National Institutes of Health, Baltimore City Hospitals, Baltimore, Maryland 21224.

INTRODUCTION

Self-perceptions of health are key components of health maintenance, since they influence efforts at self-medication as well as decisions to seek medical treatment. Physicians rely heavily on medical histories and symptom descriptions provided by patients in reaching diagnoses, and social gerontologists often employ self-ratings of health as predictors in investigations of the quality of life. All of these uses presuppose some degree of correspondence between subjective evaluation of health and objective physical status.

Most of the literature in the field of aging has employed global assessments of health, and studies have typically found that these self-ratings are moderately correlated with medical determinations of health (LaRue *et al.*, 1979; Maddox and Douglas, 1973; Tissue, 1972). These studies have also consistently shown that global health ratings are related to psychological characteristics such as health attitudes (Monroe *et al.*, 1965), morale (Friedsam and Martin, 1963; Larson, 1978; Suchman *et al.*, 1958), adjustment (Blazer and Houpt, 1979), or psychological distress (Tessler and Mechanic, 1978). Despite the demonstrable increases in many kinds of illness with age, global ratings often fail to show any marked association with age (Stenback, 1964). Markides and Martin (1979), in a sample of persons 60 and over, 61% women and 70% Mexican Americans, employed a path analysis model to investigate direct and indirect causal effects of age, objective health, and sociodemographic factors on global self-rated health. The most important predictor of self-rated health was an index of physician-rated health. Age had no significant direct effect and only small indirect effects on self-rated health.

Symptom checklists such as the Cornell Medical Index (CMI) offer certain advantages over global ratings. Since they employ specific questions about conditions and symptoms, they may provide more accurate accounts of specific illnesses, less influenced by general health attitudes. Using such measures it is possible to analyze medical conditions or body systems separately and determine whether age is differentially associated with complaints in particular somatic systems. A total score can also be obtained by summing endorsements, giving a measure of overall perceived health with a potentially higher reliability than single-item ratings. It must be noted that many of the individual items on these checklists are ambiguous with regard to etiological interpretation: Reporting chest pain may indicate indigestion instead of angina. However, these responses can legitimately be regarded as indicators of *complaints* about specific symptoms, regardless of the medical basis of the complaints.

Like global ratings, checklists are known to be influenced by both objective health (Abramson, 1966; Abramson *et al.*, 1965) and psychological factors. Clinical diagnoses of hypochondriasis or neurosis have been found to be linked to extremely high endorsement of CMI physical complaints (Ryle and Hamilton, 1962). In addition, a number of other characteristics including general anxiety (McCrae *et al.*, 1976), poor marital adjustment (Hamilton *et al.*, 1962), and psychological problems (Brodman *et al.*, 1960) have been identified as correlates of higher symptom endorsement.

Despite extensive use of self-ratings or perceptions of health, little attention has been given to conceptualizing their determinants. Many of the psychological factors identified in previous research can be hypothesized to have in common a relationship to the broad domain of personality identified as neuroticism (Eysenck, 1960; Costa and McCrae, 1980b). The present study tests this hypothesis by examining the relative influences of age and neuroticism on self-perception of health or illness for several body systems and total physical complaints. In order to separate maturational effects from generational differences and time-of-measurement effects, traditional longitudinal analyses are supplemented by analyses which approximate cross- and time-sequential designs (Schaie, 1965). In order to replicate neuroticism effects, two independent measures which fall in this domain are employed.

METHOD

Subjects. Participants in the Baltimore Longitudinal Study are a community-dwelling, generally healthy group of male volunteers, 96% white, who have agreed to return for testing at intervals of from 1 to 2 years depending on their age. The majority (80%) works in or is retired from scientific, professional, or managerial positions. Almost all (93%) are high-school graduates, and 71% are college graduates; 88% were married. At the time of their first administration of the CMI, the 1038 subjects ranged in age from 20 to 97.

Some analyses employed data from a second or third administration of the CMI. Numbers of subjects in these analyses were smaller, since many subjects had not yet participated in the study for a sufficient number of years, and since some subjects died or withdrew from the study.[4] Evidence from parallel studies on the Guilford–Zimmerman Temperament Survey (GZTS) suggests that subjects who returned for second and third administrations tended to be higher than nonrepeats in Emotional Stability, Objectivity, Friendliness, and Personal Relations and lower in Ascendance (Douglas and Arenberg, 1978).

Measures. The Cornell Medical Index (CMI; Brodman *et al.*, 1949) is a self-report symptom checklist with 195 items divided into 12 somatic sections (A–L) and 6 psychiatric sections (M–R). Two of the somatic sections, I (Frequency of Illness) and J (Fatigue), had extremely low endorsement in this sample and, thus, were combined in all analyses.

The sum of the first 12 sections yields a measure of total physical complaints; the sum of the last 6 sections provides a measure of psychiatric complaints or neuroticism. Endorsement of psychiatric items is low, with about 40% of the subjects endorsing none and an additional 20% endorsing only 1 of the 51 items.

---

[4] Of those scheduled for retest, 28% failed to take the test, due to either death or dropout. These individuals were 3 years older and two points lower on Emotional Stability than were the retested subjects, but were insignificantly higher on Total Physical Complaints.

The alternate measure of neuroticism, the Emotional Stability Scale of the GZTS (Guilford *et al.*, 1976), is a 30-item scale covering evenness vs. fluctuation of mood (7 items), perseveration of ideas (6 items), composure vs. excitability (2 items), daydreaming (3 items), feelings of guilt, loneliness, or worry (3 items), and cheerfulness vs. gloominess (7 items). Two items explicitly refer to feelings of good vs. ill health. Internal consistency in the present sample was 0.86. Handbook norms give a median score of 18; however, the median in this well-adjusted sample was 22. Correlation between the reflected Emotional Stability Scale and the CMI psychiatric score is 0.52 ($N = 915, p < 0.001$) in this sample.

Procedure. Subjects first completed the CMI as a part of a larger medical history form on their first or second visit to the Gerontology Research Center. Longitudinal analyses were restricted to subjects whose second administration was between 5.0 and 8.0 years after the first and whose third administration was between 10.0 and 17.0 years after the first. Each subject was also given the standard GZTS instructions individually and completed that questionnaire during the remainder of his first or second[5] visit to the Gerontology Research Center. For each item, subjects chose "yes" or "no" or "?." Each scale consists of 30 items, but only "yes" or "no" responses contribute to the total score. A score was invalidated for any scale with more than three "?" responses, a procedure suggested by Guilford and Zimmerman (1949). Only scores from the Emotional Stability Scale were used in the present analysis.

Analyses. Four sets of repeated-measures analyses of variance were conducted, using age and neuroticism as classifying variables. In all analyses, subjects were classified as young (20–44), middle (45–51), or old (57+) on the basis of their age at the first administration. These cutoffs were chosen to equalize the cell sizes. In two of the sets of analyses, neuroticism was operationalized by the Emotional Stability Scale of the GZTS, and subjects were classified as unstable (0–22) or stable (23–30). In the other two sets of analyses, neuroticism was operationalized by the CMI psychiatric (M–R) score, and subjects were classified as high (2 or more) or low (0) in neuroticism. Subjects endorsing one item on the M–R section were omitted from these analyses. Table I shows the sample sizes used in the repeated-measures analyses for two and three administrations of the CMI.

Cross-sequential ($N = 551$) and time-sequential ($N = 637$) analyses were also conducted on first administration data. Cross- and time-sequential designs, discussed by Baltes (1968) and Schaie (1977), are quasiexperimental designs for

---

[5] Thirty-six subjects were tested on the second visit in the early part of the study, circa 1960–1964. These subjects were an average of 5 years older than the other subjects, but did not differ in Emotional Stability or Total Physical Complaints.

Table I. Sample Sizes for Repeated-Measures Analyses

	Age group		
	20–44 (*N*)	45–56 (*N*)	56–94 (*N*)
CMI psychiatric cutoff scores			
Zero items endorsed			
Two administrations	63	43	61
Three administrations	21	21	16
Two or more items			
Two administrations	45	51	40
Three administrations	17	21	10
GZTS emotional stability scores			
23–30 (stable)			
Two administrations	60	48	58
Three administrations	22	20	18
0–22 (unstable)			
Two administrations	66	61	52
Three administrations	22	25	11
Intervals			
Between 1st and 2nd administrations: 5 to 8 years			
Between 1st and 3rd administrations: 10 to 17 years			

the study of developmental phenomena. Cross-sectional designs confound maturation with generational differences, and longitudinal designs confound maturational differences with cultural changes during the course of the study (and with the effects of repeated exposure to the test). Although they are not capable of conclusively attributing effects to aging, cohort, or time or measurement (Adam, 1978), cross- and time-sequential designs do provide an additional kind of evidence on which to base inferences. In the present study, they also permit longitudinal analysis of first administration responses from a much larger sample than that available for repeated-measures analyses. Since only first administration data are required, these analyses are not biased by selective attrition effects.

In cross-sequential analyses, independent samples of individuals born in the same historical time period are compared at different times of testing. Since recruitment into the BLSA was continuous, the present study contrasted two successive intervals of testing (1958–1963 vs. 1964–1969) rather than two distinct time points and, thus, only approximates a true cross-sequential design. Birth cohorts were defined in 6-year intervals (from 1896–1901 to 1926–1931) to match the 6-year period between times of testing.

In time-sequential analyses, independent samples of individuals of the same age are compared at different times of measurement. In these analyses, times were again defined as the intervals 1958–1963 and 1964–1969. Age groups were defined in 6-year intervals (from 26–31 to 68–73).

Studies which involve analysis of many dependent variables may capitalize on error and falsely reject the null hypothesis in some of the "significant" results. While it is possible to compensate for this through the adoption of more stringent alpha-levels or through the use of such procedures as multivariate analysis of variance, a different strategy was adopted here. By using different operationalizations of Neuroticism, different subsamples for longitudinal analyses, and a variety of analytic designs, the present paper contains what may be viewed as several replicated studies. Effects which are replicated across several such different analyses may safely be considered nonchance.

RESULTS

Table II summarizes significant ($p < 0.05$) main effects from the four sets of repeated-measures analyses. Most results are replicated in at least two analyses. All sections show an increase with higher levels of neuroticism, and three sections (Cardiovascular, Digestive, and Frequency of Illness and Fatigue), as well as the Total, show this effect in all four analyses. Longitudinal increases are shown for Sensory Systems and Genitourinary sections in all analyses, and longitudinal effects are also seen for Cardiovascular complaints and Total, which increase, and for Poor Health Habits, which decrease. Each longitudinal finding is replicated in at least one cross-sectional analysis. In addition, cross-sectional increases in Miscellaneous Diseases and Musculoskeletal complaints were found, and a cross-sectional decrease in Neurological complaints was observed in one of the four analyses.

Table III summarizes mean changes over the first interval and mean differences between emotionally stable and unstable individuals for one of the four

Table II. Summary of Main Effects from Four Repeated-Measures Analyses

	Aging		
	Cross sectional	Longitudinal	Neuroticism
Sensory Systems	Increases (3)[a]	Increases (3)	Increases (1)
Respiratory	—	—	Increases (1)
Cardiovascular	Increases (1)	Increases (2)	Increases (3)
Digestive	—	—	Increases (3)
Musculoskeletal	Increases (1)	—	Increases (2)
Skin	—	—	Increases (2)
Neurological	Decreases	—	Increases (2)
Genitourinary	Increases (3)	Increases (3)	Increases (2)
Frequency of Illness and Fatigue	—	—	Increases (3)
Miscellaneous Diseases	Increases (1)		Increases (2)
Health Habits	Decreases (1)	Decreases (1)	Increases (1)
Total	Increases	Increases (1)	Increases (3)

[a] Number of replications in parentheses.

Table III. Mean Levels of CMI Sections at Two Times and for Emotionally Stable vs. Unstable Subjects

	Time 1 vs. Time 2		Emotionally stable vs. unstable	
	50.1[a]	56.6	54.7	52.7
	($N = 345$)		($N = 166$)	($N = 179$)
Sensory Systems	1.49	1.68[b]	1.47	1.69[b]
Respiratory	1.24	1.24	0.94	1.52[b]
Cardiovascular	0.94	1.05	0.76	1.23[b]
Digestive	2.02	2.00	1.65	2.37[b]
Musculoskeletal	0.26	0.32	0.22	0.36[c]
Skin	0.41	0.43	0.32	0.52[c]
Neurological	1.14	1.05	0.93	1.26[d]
Genitourinary	1.10	1.37[b]	1.03	1.44[b]
Frequency of Illness and Fatigue	0.25	0.26	0.09	0.42[b]
Miscellaneous Diseases	1.51	1.55	1.38	1.68[d]
Health Habits	1.27	1.17	1.01	1.42[b]
Total	11.62	12.12	9.83	13.91[b]

[a] Mean age.
[b] $p < 0.001$.
[c] $p < 0.05$.
[d] $p < 0.01$.

sets of analyses in order to show the magnitude of the effects. In general, these are modest, with less than one-quarter item increase in 6 years on the two sections (Sensory Systems and Genitourinary) which show significant effects in this analysis. Effects for neuroticism are somewhat larger, with total scores about four items (42%) higher for unstable than for stable subjects.

Two significant age-group-by-time interactions were replicated. Cardiovascular complaints accelerated with age, showing the greatest increases among the oldest subjects. Poor Health Habits declined, but primarily in the old and middle groups, with little or no change in the young group.

Table IV summarizes main effects from the cross-sequential and time-sequential analyses. In cross-sequential analyses, the birth cohort factor confounds cohort and aging, and will be referred to as cohort/aging. The time factor confounds secular changes during the time of measurement and aging, and will be referred to as time/aging. Significant increases in both of these were seen for Sensory Systems, Cardiovascular, and Genitourinary complaints, as well as Total. These are most parsimoniously interpreted as aging or maturational effects. A time/aging effect was also seen for Miscellaneous Diseases, as well as a cohort/aging effect for Musculoskeletal problems. There were no significant interactions.

Table IV. Summary of Main Effects from Cross- and Time-Sequential Analyses

	Cross- sequential		Time- sequential	
	Time/aging	Cohort/aging	Time/cohort	Aging/cohort
Sensory Systems	Increases[a]	Increases[a]	—	Increases[a]
Respiratory	—	—	—	—
Cardiovascular	Increases[a]	Increases[a]	—	Increases[a]
Digestive	—	—	—	—
Musculoskeletal	—	Increases	—	Increases
Skin	—	—	—	—
Neurological	—	—	—	Decreases
Genitourinary	Increases[a]	Increases[a]	—	Increases[a]
Frequency of Illness and Fatigue	—	—	—	—
Miscellaneous Diseases	Increases[a]	—	—	Increases[a]
Health Habits	—	—	—	Decreases[a]
Total physical complaints	Increases	Increases		Increases

[a] Replicated in categorical analysis.

In time-sequential analyses, the effects can be identified as aging/cohort and time/cohort, since birth cohort is confounded with each of the other two factors. Significant increases were observed for Sensory Systems, Cardiovascular, Musculoskeletal, Genitourinary, and Miscellaneous Diseases, as well as total complaints on the aging/cohort factor. Significant decreases on this factor were seen for the Neurological and Poor Health Habits sections. By contrast, *no* effects proved significant on the time/cohort factor, nor were any interactions significant. This is strong evidence that maturational rather than generational differences or cultural changes during this period of measurement are responsible for the observed effects. It should also be noted that these analyses generally confirm the repeated-measures analyses; magnitudes of the score differences are also comparable.

Supplementary Categorical Analysis. Because of low endorsement frequencies, distributions for most of the section scores were skewed. Although analysis of variance is known as a robust technique, relatively insensitive to departures from normality (McNemar, 1962), there was some concern that the results might be distorted. As an alternative, the Grizzle–Starmer–Koch (GSK) approach to the analysis of categorical data was considered (Grizzle *et al.*, 1969). This technique was designed to analyze categorical or nominal data, although it can be applied to ordinal data, such as the number of items endorsed. However, in order to test hypotheses, there should be at least 10 observations in every cell (Kleinbaum and Kupper, 1978). To fit the present data to this requirement, it would be necessary to collapse scores into two categories, e.g., "none" vs. "some." Information is lost in this process, and there is room for some subjectivity in the choice of a cutoff point. Nevertheless, as a check on the analysis of variance results, categorical analyses were performed on dichotomized section scores to parallel the cross- and time-sequential analyses. Scores for Sensory Systems, Digestive, Neurological, Miscellaneous Diseases, and Health Habits were classified as "none or one" vs. "two or more" endorsements; scores on all other sections were dichotomized as "none" vs. "one or more" endorsements.

The Funcat procedure of the Statistical Analysis System (SAS; Helwig and Council, 1979) was employed to analyze responses. Times of measurement and birth cohorts (as defined in the cross-sequential analyses) were used as factors in one set of analyses, and times of measurement and age groups (as defined in the time-sequential analyses) were factors in the second. Dichotomized section scores were treated as responses. In the first set of analyses, significant ($p < 0.05$) main effects were found for Sensory Systems, Cardiovascular, Genitourinary, and Miscellaneous Diseases. Significant effects for cohort/aging were found for Sensory Systems, Cardiovascular, and Genitourinary sections.

In the second set of analyses corresponding to the time-sequential analyses of variance, no significant effects were observed for time/cohort. Aging/cohort effects were found for Sensory Systems, Genitourinary, Miscellaneous Diseases, and Health Habits. These effects are indicated in Table IV.

Examination of Table IV will show that each main effect in the categorical analysis was also found in the analysis of variance, although the analysis of variance suggested some effects not replicated or found by the categorical analysis. The greater sensitivity of the ANOVA procedure may be due to the additional information available from using scores as a continuous variable. In any case, the general pattern of results is similar from the two sets of analyses.

DISCUSSION

The data presented here show a consistent pattern across different methods of analysis and different operationalizations of neuroticism. Age has a selective effect on physical complaints, while neuroticism appears to produce a more general and diffuse effect on physical complaints. The fact that significant effects were found despite the restriction of range in this healthy and well-adjusted sample argues that the latter relationship must be quite strong, and that the influence of neuroticism on health perception might be even more pronounced in the general population.

The specific effects associated with aging are not surprising. The prevalence of sensory, cardiovascular, musculoskeletal, and genitourinary problems in this age group is well-known (Shanas and Maddox, 1976). It is somewhat puzzling that respiratory and digestive complaints do not appear to increase with age; this may be an artifact of the particularly healthy sample used. The fact that only certain systems show age-related increases may account in part for the fact that global health ratings are only weakly related to age.

Corroborating evidence that older people do not present large numbers of complaints is provided by Eckstein (1978). As a full-time geriatric physician, Eckstein comments that, "given the large burden of illness and disabilities the elderly endure . . . the complaints of the elderly are remarkably low-keyed and valid" (p. 16). He attributes this to the unwillingness of most older patients to fit the pejorative stereotypes of old complainers. In addition, it can be speculated that health expectations decline with age. Conditions like fatigue after exercise, considered medically significant by younger people, may be regarded as a normal part of aging by the elderly and, thus, not reported as a complaint. Finally, well-adjusted older individuals may show a realistic concern for their health without a greater number of complaints. Most older people, for example, learn to be more cautious in walking in order to avoid falls.

What we do not see in the present sample is any increase in unrealistic, unproductive obsession with health and bodily functioning. These findings contradict the stereotype of aging people as hypochondriacs or "crocks" (Butler, 1978), but they are consistent with other literature. Using psychodynamically oriented psychiatric ratings of hypochondriasis, Gianturco and Busse (1978) report a longitudinal decline in hypochondriasis. In a sample of 176 community-dwelling men and women aged 60–90, 6% became more hypochondriacal, whereas 26% became less so. This does not mean that there are no elderly hypochondriacs; however, the proportion of such people is no higher in old age than in youth or middle age (cf. Shanas and Maddox, 1976).

At any age, excessive complaints are associated with neuroticism, or poor psychological adjustment, which is itself unrelated to age. Gianturco and Busse (1978) report that hypochondriasis was associated with lower levels of happiness, fewer friends, and depression and that "the vast majority of the severe neurotics were . . . hypochondriacs" (p. 12). Luborsky *et al.*, (1973), in a review of over 50 studies in the field of psychosomatic medicine, found a variety of psychological factors including resentment, frustration, depression, anxiety, and helplessness associated with a range of ailments from "cold hands" to cancer; there was no evidence of symptom specificity. Similarly, Tessler and Mechanic (1978) found that psychological distress, whether measured as negative affect, nervousness, or global unhappiness, was associated with lower self-ratings of health. Together with the present findings, these studies lead to a clear conclusion. Any manifestation of neuroticism—hostility, depression, anxiety, vulnerability to stress—is likely to be associated with diffuse somatic complaints.

The nature of the association between neuroticism and perceived health is not fully understood. There is some evidence that more neurotic individuals do not suffer from a higher incidence of fatal illness (Keehn *et al.*, 1974), but it is possible that they are more frequently troubled by minor health problems, particularly such psychosomatic symptoms as fatigue, gastrointestinal problems, or palpitations. Luborsky *et al.* (1973) suggest that psychological variables may be involved in determining the onset or exacerbation of an illness to which an individual is predisposed. The range of complaints associated with neuroticism might suggest that individuals higher in neuroticism are simply more likely to complain. However, Meadow *et al.* (1978) have provided evidence that perception of somatic functioning, particularly autonomic frequency and reactivity, is significanlty related to neuroticism (stress reactivity). This evidence is consistent with the interpretation that individuals higher in neuroticism are more sensitive or attentive to bodily states.

In any case, it is clear that clinicians or researchers who employ self-reports of health or illness should be aware of the pervasive role of neuroticism. LaRue *et al.* (1979) have suggested that self-ratings of health "could provide a valid cost-effective measure of health assessment" (p. 687). But to the extent that self-ratings of health share the same determinants as measures of somatic complaints, such ratings will be determined by both objective health and neuroticism. Research concerned with the influence of physical health on morale, sick-role behavior, or adjustment should either use objective measures of health or supplement self-ratings with measures of neuroticism in order to control for its effects.

REFERENCES

Abramson, J. H. (1966). The Cornell medical index as an epidemiological tool. *Am. J. Public Health:* 287-298.

Abramson, J. H., Terespolsky, L., Brook, J. G., and Kark, S. L. (1965). Cornell medical index as a health measure in epidemiological studies: A test of the validity of a health questionnaire. *Br. J. Prevent. Med.* 19: 103-110.

Adam, J. (1978). Sequential strategies and the separation of age, cohort, and time-measurement contributions to developmental data. *Psychol. Bull.* 85: 1309-1316.

Baltes, P. B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human. Develop.* 11: 145-171.

Blazer, D. G., and Houpt, J. L. (1979). Perception of poor health in the healthy older adult. *J. Am. Geriatr. Soc.* 27: 330-334.

Brodman, K., Erdmann, A. J., Jr., Lorge, I., and Wolff, H. G. (1949). The Cornell medical index: An adjunct to medical interview. *J. Am. Med. Assoc.* 140: 530-534.

Brodman, K., Erdmann, A. J., and Wolff, H. G. (1960). *The Cornell Medical Index – Health Questionnaire Manual*, Cornell Univesity Medical College, Ithaca, N.Y.

Butler, R. N. (1978). The doctor and the aged patient. In Reichel, W. (ed.), *The Geriatric Patient*, HP Publishing, New York.

Costa, P. T., Jr., and McCrae, R. R. (1980a). The influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *J. Personal. Soc. Psychol.* 38: 668-678.

Costa, P. T., Jr., and McCrae, R. R. (1980b). Still stable after all these years: Personality as a key to some issues in aging. In Baltes, P. B., Brim, O. G. (eds.), *Life-Span Development and Behavior*, Vol. III, Academic Press, New York, pp. 65-102.

Costa, P. T., Jr., McCrae, R. R., and Arenberg, D. (1980). Enduring dispositions in adult males. *J. Personal. Soc. Psychol.* 38: 793-800.

Douglas, K., and Arenberg, D. (1978). Age changes, cohort differences, and cultural change on the Guilford-Zimmerman temperament survey. *J. Gerontol.* 33: 737-747.

Eckstein, D. (1978). Common complaints of the elderly. In Reichel, W. (ed.), *The Geriatric Patient.* HP Publishing, New York.

Eysenck, H. J. (1960). *The Structure of Human Personality.* Methuen, London.

Friedsam, H. J., and Martin, H. W. (1963). A comparison of self and physicians' health ratings in an older population. *J. Health Human Behav.* 4: 179-183.

Gianturco, D. T., and Busse, E. W. (1978). Psychiatric problems encountered during a long-term study of normal aging volunteers. In Issacs, A. D., and Post, F. (eds.), *Studies in Geriatric Psychiatry.* Wiley, New York.

Grizzle, J. E., Starmer, C. F., and Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics.* 25: 489-504.

Guilford, J. P., and Zimmerman, W. S. (1949). *The Guilford-Zimmerman Temperament Survey: Manual of Instructions and Interpretations.* Sheridan Supply Co., Beverly Hills, Calif.

Guilford, J. S., Zimmerman, W. S., and Guilford, J. P. (1976). *The Guilford-Zimmerman Temperament Survey Handbook: Twenty-Five Years of Research and Application,* Edits, San Diego, Calif.

Hamilton, M., Pond, D. A., and Ryle, A. (1962). Relation of CMI responses to some social and psychological factors. *J. Psychosom. Res.* 6: 157-165.

Helwig, J. T., and Council, K. A. (1979). *SAS Users Guide, 1979 Edition.* SAS Institute, Inc., Raleigh, N.C.

Keehn, R. J., Goldberg, I. D., and Beebe, G. W. (1974). Twenty-four year mortality follow-up of army veterans with disability separations for psychoneurosis in 1944. *Psychosom. Med.* 36: 27-46.

Kleinbaum, D. G., and Kupper, L. L. (1978). *Applied Regression Analysis and Other Multivariate Methods.* Doxbury Press, North Scituate, Mass.

Larson, R. (1978). Thirty years of research on the subjective well-being of older Americans. *J. Gerontol.* 33: 109-125.

LaRue, A., Bank, L., Jarvik, L., and Hetland, M. (1979). Health in old age: How do Physicians' ratings and self-ratings compare? *J. Gerontol.* 34: 687-691.

Luborsky, L., Docherty, J. P., and Penick, S. (1973). Onset conditions for psychosomatic symptoms: A comparative review of immediate observation with retrospective research. *Psychosom. Med.* 35: 187-204.

Maddox, G. L., and Douglas, E. B. (1973). Self-assessment of health: A longitudinal study of elderly subjects. *J. Health Soc. Behav.* 14: 87-93.

Markides, K. S., and Martin, H. W. (1979). Predicting self-rated health among the aged. *Res. Aging* 1: 97-112.

McCrae, R. R., Bartone, P. T., and Costa, P. T., Jr. (1976). Age, anxiety, and self-reported health. *In. J. Aging Human Develop.* 7: 49-58.

McNemar, Q. (1962). *Psychological Statistics.* 3rd ed., Wiley, New York.

Meadow, M. J., Kochevar, J., Tellegen, A., and Roberts, A. H. (1978). Perceived somatic response inventory: Three scales developed by factor analysis. *J. Behav. Med.* 1: 413-426.

Monroe, R. T., Whiskin, F. E., Bonacich, P., and Sewell, W. O. (1965). The CMI questionnaire as a measure of health in older people. *J. Gerontol.* 20: 18-22.

Ryle, A., and Hamilton, M. (1962). Neurosis in 50 married couples. *J. Ment. Sci.* 108: 265.

Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychol. Bull.* 64: 92-107.

Schaie, K. W. (1977). Quasi-experimental research designs in the psychology of aging. In Birren, J. E., and Schaie, K. W. (eds.), *Handbook of the Psychology of Aging.* Van Nostrand Reinhold, New York.

Shanas, E., and Maddox, G. L. (1976). Aging, health, and the organization of health resources. In Binstock, R. H., and Shanas, E. (eds.), *Handbook of Aging and the Social Sciences,* Van Nostrand Reinhold, New York.

Stenback, A., (1964). Physical health and physical disease as objective fact and subjective experience. *Arch Gen. Psychiat.* 11: 290-301.

Suchman, E. A., Phillips, B. S., and Streib, G. F. (1958). An analysis of the validity of health questionnaires. *Soc. Forces* 36: 223-232.

Tessler, R., and Mechanic, D. (1978). Psychological distress and perceived health status. *J. Health Soc. Behav.* 19: 254-262.

Tissue, T. (1972). Another look at self-rated health among the elderly. *J. Gerontol.* 27: 91-94.

Psychological Bulletin
1982, Vol. 92, No. 1, 238-250

An Approach to the Attribution of Aging, Period, and Cohort Effects

Paul T. Costa, Jr., and Robert R. McCrae
Gerontology Research Center, Baltimore City Hospitals, Baltimore, Maryland
National Institute on Aging, National Institutes of Health, Bethesda, Maryland

In an attempt to extricate the inherently confounded factors of maturation, cultural change, and generational differences, life-span methodologists have proposed a variety of analytic designs and interpretative decision rules. Recent critiques have shown that the proposed rules are inadequate and that there are logical limits to any such set of rules. A number of alternatives have been offered—most of which require the investigator to have strong, theoretically guided hypotheses, and presume that the data conform strictly to the demands of the design. The present article addresses the common situation in which data analysis is exploratory rather than hypothesis testing and in which the model is applied to data that only loosely meet the requirements of the design. In this case, aging, period, or cohort effects can be inferred if the researcher is willing to make appropriate restrictive assumptions and uses scientific judgment rather than fixed decision rules. The application of judgmental principles is illustrated on two longitudinal data sets, and it is argued that the analytic designs are useful if intelligently applied and interpreted.

Students of human aging and of social or cultural change are faced with certain unavoidable inferential problems. Generational differences and maturational and cultural changes are inevitably confounded because each individual is born and ages in one and only one historical period. Methodologists in fields such as life-span development (Schaie, 1965) and cohort analysis (Mason, Winsborough, Mason, & Poole, 1973) have proposed various designs in an attempt to extricate the correct effects from the confounds, and these designs have been widely taught and applied (Birren & Schaie, 1977; Botwinick, 1973; Maddox & Wiley, 1976). Recently, however, critiques of the decision rules originally offered by Schaie (1965) have been made (Adam, 1978; Glenn, 1976) that cast serious doubt on the utility of any of the proposed analytic designs. Although these critiques had long been anticipated by some writers (Baltes, 1968), and although a number of partial resolutions to the problems have been offered (Baltes, Cornelius, & Nesselroade, 1979; Mason et al., 1973; Schaie & Baltes, 1975), there is still confusion in the field. A careful reading of the recent literature suggests that no single solution to the problem of confounding can be found and that the choice of interpretative principles depends on the nature of the data, the goals of the investigator, and the state of knowledge in the area.

In this article we propose to discuss interpretation of results from the general developmental model for the common situation in which the data do not conform rigorously to the design, and the analysis is essentially exploratory rather than hypothesis testing. We attempt to justify our interpretations not on strictly mathematical grounds but on broader principles of scientific inference. Within the closed system of age (or maturation), time (or period), and cohort (or generation), there are no solutions to the basic logical problem; if, however, we impose certain restrictions on the system, and if we look beyond that system to other independent sources of evidence, we can begin to establish an interpretation of the data. Within this approach, the sequential designs remain potentially important as one source of evidence. Their use should by all means be encour-

---

Requests for reprints should be sent to Paul T. Costa, Jr., Chief, Section on Stress and Coping, Gerontology Research Center, Baltimore City Hospitals, Baltimore, Maryland 21224.

aged, but the data they yield must be interpreted in the light of many additional considerations.

General Developmental Model

The central methodological problem in the field of life-span developmental psychology is the identification of maturational effects or aging changes. Because of the time spans covered by the aging process, developmental phenomena are often difficult to discern, and successive replications are often impossible. Few investigators can afford to dedicate their careers to one longitudinal study, let alone a series of studies each building on the results of the previous ones.

Even more basically, age is not an independent variable. It cannot be manipulated; subjects cannot be assigned to "aging" and "nonaging" conditions. For this reason, methodologists have been forced to develop quasi-experimental designs for the study of maturational effects (Schaie, 1977) from which to make inferences about aging. These designs are never perfect substitutes for true experiments (Campbell & Stanley, 1963), and they always allow the possibility of alternative explanations.

Cross-sectional analyses use the familiar strategy of comparing two or more age groups on a single testing occasion. For example, 20-year-olds may be compared with 80-year-olds. Cross-sectional studies can conveniently examine a wide age range, but, as gerontologists have long been aware, the observed effects may be due to generational differences rather than to maturation, and this factor in an analysis of variance (ANOVA) design should be referred to as an "aging/cohort" effect, not an "aging" effect (see Whitbourne & Waterman, 1979).

Traditional *longitudinal* designs attempt to circumvent this problem by following the same group of individuals over two or more times of measurement (and thus at two or more ages). Age-related effects in such a design cannot be attributed to generational differences because the individual's date of birth never changes. In fact, the subjects can be regarded as their own controls, matched not only on generation but also on a host of other variables. In analyses of longitudinal data, repeated measures ANOVAS (Winer, 1971) can be used, giving greater statistical power. Traditional longitudinal designs, however, are also subject to alternate interpretations. Changes that occur between the first measurement and the second may be due to intervening historical events (time) rather than to aging, and for some tests, previous exposure or practice may be responsible for effects. Thus, the classifying factor here should be considered a "time/aging/practice" effect.

Schaie (1965) proposed some additional designs that he believed would be useful in avoiding these problems. In the *cross-sequential* design,[1] independent samples of individuals born in the same period are compared at different times of measurement. Because a given individual is measured only once, practice effects are eliminated. In setting up an analysis of variance for cross-sequential data, birth cohort and times of measurement are used as classifying factors. Aging, however, is confounded with both of these effects, and, for purposes of clarity, the classifying factors can be labeled "cohort/aging" and "time/aging."

In the *time-sequential* design, independent samples of individuals of a particular age are compared at different times of measurement. Age group and time of measurement are separated, but both are confounded with cohort, and the factors are identifiable as "aging/cohort" and "time/cohort."

The crucial point in Schaie's (1965) original approach was that, although each of these designs is ambiguous in interpretation when used alone, it may be possible to distinguish age, period, and cohort effects if all are used and analyzed simultaneously. Schaie's model and its elaborations by other writers have greatly affected the course of research in the field of life-span development, and cross- and time-sequential designs have been widely applauded and occasionally used. Two criticisms, however, have arisen that have shaken confidence in the use of these longitudinal designs.

---

[1] We are here following what we perceive to be the convention in use in the field (e.g., Douglas & Arenberg, 1978; Whitbourne & Waterman, 1979) in identifying these designs. Technically, what we have called the traditional longitudinal design is considered a cross-sequential design by Schaie (1977), insofar as individuals born in the same cohort are measured at different times. Nevertheless, we restrict the term *cross-sequential* to designs in which independent samples born in the same period are compared at different times of measurement.

First, Schaie's decision rules have been criticized as being insufficiently restrictive. Within the set of assumptions that Schaie (1965) proposed, alternative and in fact radically different interpretations of the same data are possible. Only more stringent assumptions can allow unequivocal inference. Second, theorists in the life-span tradition have pointed out numerous reasons for believing that Schaie's rules are too restrictive to fit the facts. Caught between these two criticisms, the utility of the general developmental model seems directly threatened.

Logical problems. The first criticism is logical in nature and arises from the fact that these designs always involve more unknowns than there are equations from which to solve for them. Given a knowledge of age and of time of measurement, cohort is always exactly fixed, and the same holds for any two other variables in the set. Originally, Schaie (1965) argued that this mutual confounding could be deconfounded by the simultaneous use of several different designs, and he offered numerical decision rules based on *F* ratios that he felt could be used to determine which effects were responsible for observed differences. Adam (1978), however, demonstrated mathematically that these decision rules were unsound. In essence, it is always possible to explain an observed difference in more than one way. For example, a pure maturational effect that was linear across the life span should be seen in significant effects for cohort and time in the cross-sequential design and for age but not time in the time-sequential design. An identical pattern of results, however, could be obtained if linear effects for cohort and for time of measurement were present in equal amounts but in opposite directions. There are other possible sets of effects that would also produce these results. No mathematical manipulation of these data can ever deconfound these effects under the assumptions imposed by Schaie.

Substantive problems. Schaie's (1965) decision rules and Adam's (1978) criticisms apply only to a simplified model in which effects for all three variables are assumed to be linear and there are no interactions. There are, however, substantive reasons to believe that these assumptions are not necessarily valid, leading to a second criticism. Economic conditions and child-rearing patterns that may be responsible for generational differences are often cyclical rather than linear. Some cognitive functions are believed to show an accelerated decline with age. The possibilities of interactions are legion: Will individuals reared in the Depression react to current economic instability in the same way as those reared in the relatively more prosperous times before and after?

Consider another set of issues. What is the appropriate unit for grouping subjects? When does a cohort start and end? In 5-year intervals or in 10-year intervals? What constitutes a meaningful "age group"? Should one class subjects by decades or by 7-year cycles or by role-defined periods? What are appropriate intervals at which to conduct data collection? Would time-of-measurement effects be obscured or enhanced by continuous measurement as opposed to measurement every few years?

All of these problems may easily lead the researcher to a nihilistic position, perhaps most forcefully stated by Gergen (1977), who argued that a "chance" model of change in adulthood is necessary and that the knowledge that gerontologists of this period in history generate may have no relevance to succeeding generations.

Nevertheless, scientists have always had to face the philosophical critique that inductive knowledge is never certain. They have learned that it is useful to make simplifying assumptions and empirical generalizations and that complex possibilities should be ignored until the inadequacies of simpler approaches become evident; also, they have developed a body of conventions that is useful in interpreting the fundamentally ambiguous results of any scientific investigation. Among these principles are replicability, parsimony, and consistency. The application of these principles to a particular body of data is occasionally formalized (see, e.g., Rosenthal, 1978, on the quantification of replicability), but for the most part they require the use of what is called *scientific judgment*. Adam (1978) showed that rigid application of decision rules is insufficient for separating maturational effects from other possible causes of change in longitudinal studies; the present article argues that the application of scientific judgment may be more fruitful.

Alternative Restrictive Assumptions

Given the fact that Schaie's (1965) assumptions cannot sustain the unambiguous attributions of cause to any of the three effects, or combination of effects, inference can be made only by imposing more restrictive assumptions. Baltes (1968) pointed out some time ago that the three designs—cross-sequential, time-sequential, and cohort-sequential—are redundant. The use of any two exhausts the independent information, as Adam (1978) also demonstrated.[2] Baltes proposed that both time- and cross-sequential designs be dropped in favor of cohort-sequential designs. In these designs, time of measurement is confounded with age on one factor and with birth cohort on another. Baltes argued that time of measurement is unlikely to influence the variables of interest to developmentalists and can be safely eliminated a priori as a source of variation. The cohort-sequential design under this assumption then becomes a simple test of the independent effects of age and cohort. By an extension of this logic, researchers who are willing to rule out cohort effects, a priori, could use the time-sequential design; those who are willing to rule out maturation effects could use the cross-sequential design.

In practice, the cohort-sequential design is practicable only when very short time intervals are used, and it may be most useful for studies of child development. In longitudinal studies of adults, where there is a wide range of ages and birth cohorts but a small range of times of measurement, it may be more reasonable to use cross- and time-sequential strategies. Because these together are formally equivalent to the cohort-sequential design, the same inferences can be drawn. That is, if one can rule out times-of-measurement effects, one can infer age and cohort effects from the combined use of these two designs

Of course, the major drawback to this approach is the fact that the assumption that there are no time effects (or age or cohort effects) may be wrong. So restrictive an approach does not allow for any falsification of the initial assumption. A more wide-ranging strategy combining repeated measures longitudinal analysis along with sequential designs might give a better basis for inferring effects. In addition, in many instances longitudinal investigations are exploratory, and we have neither desire nor rationale to rule out a particular category of causes.

An alternative that uses multiple regression has been proposed by Mason et al. (1973). To obtain a solution to the regression equations, the researcher must assume that some pair of levels of one of the variables are equal, say, that the two oldest groups do not differ on the variable. More informative results can be obtained by setting additional levels equal. This strategy is appealing when very specific hypotheses can be made. Instead of an entire class of effects being eliminated, as the Baltes strategy requires, only certain levels of effects need be restricted. On the other hand, this strategy, too, has problems. Table 5 in the Mason et al. article (p. 251), for example, shows statistically significant age effects in data that were artificially constructed to represent pure cohort effects, and the authors admit that confounding would be worse if the "pure" effects were linear rather than nonlinear. The authors caution that three-way analysis "is difficult unless the researcher entertains relatively strong hypotheses about the nature of aging, period, and cohort effects" (p. 243).

The alternative used in the present article relies on the analysis of variance method proposed by Schaie (1965), and, following him, it assumes that effects are monotonic. We make the additional assumption that one and only one effect is operating on each variable, and our problem is determining to which one, if any, of the effects we should attribute the results. We choose to entertain the possibility that any of the three may be responsible for observed effects, but we do not specify which it is. We eliminate the possible but less parsimonious explanation that two or three of the factors are operating jointly.

---

[2] In practice, however, analyses are usually conducted on subsets of individuals who fit within the chosen levels of the classifying factor. When time of measurement is crossed with cohort (in cross-sequential designs), the oldest and youngest subjects may be omitted, although they could be included in time-sequential designs. Because subjects differ somewhat, the information from different analyses is not perfectly redundant and may be useful in considering internal consistency of results.

We offer this alternative despite its restrictive assumptions because there appear to be many areas in which very little guidance from theory or research is applicable. Child development is far better understood today than adult development, primarily because the span of time required for research is so much shorter. Psychologists have been able to observe the sequences of cognitive development, for example, in many different generations of children and under many different educational and historical influences. With so much knowledge, the researcher can easily rule out certain sources of variance, probably including time of measurement, in analyzing data. In contrast, we have only begun to collect longitudinal data on adults in most areas, and the influences of time of measurement, maturation, and birth cohort on variables like well-being and perceived health are largely unknown. Theoreticians have multiplied the number of factors that might influence results, but they have not offered a clear basis for predictions. In these circumstances, researchers must begin basic explorations, and the restrictive assumptions adopted here are intended to guide such early efforts.

Practical Problems in Real Data Sets

Like most statistical problems, Schaie's general developmental model has usually been discussed in the abstract. Ideal experimental designs are posited and hypothetical effects considered. In real situations, the Schaie designs are often applied to data that fall far short of the ideal. In these cases, effects are observed that may not be predicted from the model. There must be room in interpreting the data for chance results, and the researcher must know when to ignore significant but spurious results and when to interpret nonsignificant trends. All of these problems require an exercise of judgment that simple decision rules cannot replace.

Practically speaking, it has rarely been possible to analyze existing longitudinal data within the Schaie framework, and it is expensive and time consuming to institute new studies. In the ideal design, repeated longitudinal measurements would be supplemented by independent samples drawn every few years over a period of time long enough to ensure adequate sampling of times of measurement. Sampling procedures would need to be identical, or effects could be attributed to changes in sample composition rather than to aging or time of measurement.[3] Whitbourne and Waterman (1979), for example, devoted considerable time to a discussion of the comparability of the independent samples they use to deal with this problem. In the use of existing data, other researchers may have to acknowledge that their successive samples are *not* comparable in some respects. Under these conditions, the mechanical application of decision rules is unwise. Instead, sound judgment about the likely effects of deviations from the ideal design are most likely to result in accurate and replicable interpretations of the data.

Another respect in which real data usually depart from the ideal is in the number of levels of measurement for each of the three factors. In the study of adult development it is easy to sample a wide age range and a wide range of birth cohorts, but sampling an equally wide range of times of measurement is usually impractical. Not only does this limit the generalizability of findings regarding period effects; it also presents statistical complications (Botwinick & Arenberg, 1976). A small or modest aging effect may be detectable across the full adult life span, but a period effect of equal size (in terms of effect per year) may not reach significance in an interval of only 5 or 10 years. In the data to be presented, the failure to find any time-of-measurement effects may be attributable to the restricted range of this source of variation or to the lack of sufficient statistical power in the designs. For developmentalists, however, period effects themselves are largely nuisance variables of little intrinsic interest. If they are large enough to cause interpretative problems, they should appear even when only a restricted range of times is sampled; if they are very small, then although a real effect may be missed, that loss will be of little practical or theoretical import. In contrast to the solution proposed by Baltes et al. (1979), in which period effects are ruled out a priori, the present plan allows the detection at least of large period effects if they do in fact exist.

---

[3] This presents a special problem when older persons are studied because death can alter the population being sampled. This particular problem has been addressed (Baltes, Reese, & Nesselroade, 1977, Chapters 5 and 6) by limiting analysis of the first time data to those subjects who have survived at the time of the second round. Unfortunately, this entails tracing the original subjects, a problem that it was hoped the use of cross-sequential designs would obviate.

The recognition that judgment is needed in the interpretation of developmental designs is by no means original. Baltes et al. (1977), in particular, urged a distinction between the descriptive and the explanatory function of such methodologies and argued that we need to look elsewhere for causal explanations of effect, and Glenn (1981) recently renewed the call for judgment in interpreting cohort data.

An Example: Longitudinal Changes in the Chicago Attitude Inventory

Data collection began in 1958 for the Baltimore Longitudinal Study of Aging (BLSA), an ongoing intramural project of the National Institute on Aging conducted at the Gerontology Research Center. At that time, the Chicago Attitude Inventory (CAI; Havighurst, 1951) was one of the few available instruments tailored for the study of personal adjustment among the aged. Together with the Chicago Activities Inventory, it was incorporated into a questionnaire titled "Your Activities and Attitudes" (Burgess, Cavan, & Havighurst, 1948) that has been administered at the BLSA for 20 years (Stone & Norris, 1966). The questionnaire was intended originally for respondents over the age of 55 but was given to all participants to allow cross-sectional comparisons. All of the items and scales analyzed in the present study are applicable to adults of any age.

The present study uses cross-sequential and time-sequential as well as traditional longitudinal designs to examine stability or change in these measures and to illustrate the interpretative principles that permit inferences about aging, time, and cohort effects.

Method

Participants in the BLSA are a community-dwelling, generally healthy group of male volunteers, 96% white, who have agreed to return for testing at fixed intervals.[4] The majority (80%) work in or are retired from scientific, professional, or managerial positions. Almost all (93%) are high school graduates, and 71% are college graduates; 88% are married. At the time of their first administration of the CAI, subjects ranged in age from 17 to 97. As in all longitudinal studies, there has been some attrition. Subjects who remained in the study tended to be psychologically better adjusted than those who dropped out (Douglas & Arenberg, 1978).

The eight sections of the CAI each consist of seven items to which the respondent marked "Agree," "Disagree," or "?" Items were coded −1 (poor attitude), 0 (?), or +1 (good attitude) and were summed to form eight scales. This method of scoring follows the original procedure (Cavan, Burgess, Havighurst, & Goldhammer, 1949, p. 119).

The CAI was supplemented by two global items taken from the Activities section of the "Your Activities and Attitudes" questionnaire. Item L6, Assessment of Life, asks the respondent, "As you look back over your life, in general would you call it very happy, moderately happy, average, or unhappy?" Because less than 1% of the sample checked "unhappy," this response category was combined with "average" in all analyses. Item L8, Satisfaction with Accomplishments, asks subjects if they feel "well-satisfied, reasonably satisfied, or dissatisfied" with what they have accomplished in life.

Each subject was given the CAI individually with standard instructions and completed the questionnaire during the remainder of his 3-day visit to the Gerontology Research Center. Participants over 70 years of age are on a yearly visit schedule; those 60–70 return every 18 months; and those under 60 return every two years. The CAI was given on the first or second visit and readministered on the fifth and ninth visits.

Cross-sequential analyses. Because recruitment into the BLSA was continuous, the present study contrasted two successive intervals of testing (1958–1963 vs. 1964–1969) rather than two distinct time points and thus only approximates a true cross-sequential design. Birth cohorts were defined in 6-year intervals (from 1896–1901 to 1926–1931) to match the 6-year period between times of testing. The design is depicted in Table 1.

Analyses were conducted on a sample ($N = 425$) including all subjects who had first-administration data

---

[4] Female participants entered the study in 1978, and there are currently about 300 women enrolled. Note, however, that the analyses reported here use data from men only.

Table 1
ANOVA Design and Cell Frequencies for Cross- and Time-Sequential Studies

Group	Time of measurement (time/aging) 1958–1963	1964–1969
Cross-sequential[a]		
Birth cohorts (cohort/aging)		
1926–1931	29	20
1920–1925	40	45
1914–1919	45	41
1908–1913	54	33
1902–1907	37	28
1896–1901	31	22
Time-sequential[b]		
Age groups (aging/cohort)		
32–37	39	17
38–43	38	36
44–49	54	41
50–55	48	40
56–61	38	31
62–67	24	19
68–73	29	23

[a] $N = 425$. [b] $N = 477$.

prior to 1970. Insofar as 10 dependent variables were examined, an overall multivariate analysis of variance (MANOVA) was performed first to guard against spurious effects.

Time-sequential analyses. In these analyses, times of measurement were again defined as the intervals 1958–1963 and 1964–1969. Age groups were defined in 6-year intervals (from 32–37 to 68–73) to match the 6-year period between times of testing. Data from 477 subjects tested before 1970 were analyzed, and MANOVAS were used again. The design for these analyses is also given in Table 1.

Repeated measures analyses. Traditional cross-sectional and longitudinal analyses were combined in repeated measures ANOVAS using age groups as a second factor (see Table 2). As in the time-sequential analyses, age groups were defined in 6-year intervals according to age at first administration.

Because the time interval between administration of the CAI varied, analyses were restricted to subjects who completed the second CAI between 4 and 8 years after the first and the third CAI between 4 and 8 years after the second. The average time interval was approximately 6 years.

In the first set of analyses, data from 239 subjects who had taken the CAI at least twice were examined, using two levels of the repeated factor. In the second set of repeated measures, data from 76 subjects who had taken the CAI at least three times were examined in a design using three levels of the repeated factor. Numbers of subjects in these analyses were smaller than in cross- and time-sequential analyses because many subjects had not yet participated in the study a sufficient number of years to be retested, and some had withdrawn or died.

Interpreting Results

Statistically significant ($p < .05$) main effects from all of these analyses are summarized in Table 3. There were no significant interactions. From this array of findings it is necessary to reach some conclusions, to make inferences based on the data. Insofar as the purpose of this article is to illustrate the process of interpretation as applied to developmental analyses, we spell out at some length the principles we apply and their rationale. Our conclusions are that (a) usefulness and work show an age-related decline but not necessarily a maturational decline and (b) personal adjustment is generally quite stable across the adult years. Our logic is based on the following principles.

Parsimony. We make certain simplifying assumptions in considering the data from cross- and time-sequential designs. In choosing the intervals for age, time of measurement, and cohort and in using continuous rather than discrete age and time units, we

Table 2
ANOVA Designs and Cell Frequencies for Two- and Three-Time Repeated Measures

Age group (aging/cohorts)	Administration (time/aging/practice)
Two measurements[a]	
32–37	25
38–43	54
44–49	45
50–55	44
56–61	28
62–67	23
68–73	20
Three measurements[b]	
32–37	8
38–43	19
44–49	13
50–55	15
56–61	6
62–67	7
68–73	8

Note. The same subjects were measured at each administration.
[a] $N = 239$. [b] $N = 76$.

assumed that the effects we were looking for would be linear. We had no justification for a more complex model, and the data seem to be generally consistent with these assumptions. Thus, when an effect is listed in Table 3 as "increased," it should be understood that there is a monotonic increase across all levels of the factor and that the overall F value is significant. The fact that no interactions were significant also supports the assumptions made. It is logically possible that different results would have been obtained if some different set of age or generation cuts had been made, but it is more parsimonious to disregard that possibility in the present study.

We also assume that any effect that we observe is due to a single cause rather than to a compound cause. It is true, as Adam (1978) showed, that what appears to be an age effect could result from the equal and opposite effects of time of measurement and generational change. For example, it may be that as people age they lose their sense of usefulness. It is also possible that usefulness is unrelated to maturation; perhaps over the past 90 years each succeeding generation has been raised with a higher sense of usefulness but that each passing year in this historical period has diminished the sense of usefulness of every individual. Both of these are possible, but we will assume that the former is the more likely because it is simpler (and because it is more consistent with our knowledge from other sources). Although Glenn (1981) gave an example in which the most parsimonious explanation is probably not the correct one, we provisionally make the assumption that any effect we observe is due to aging or to generational differences or to cultural change but not to a combination of these. The data may force us to abandon the assumptions.

Logical elimination. Under the assumption that one and only one "cause" is responsible for the observed effects, we can logically eliminate certain possibilities. The "cause" we are looking for must be a factor in all statistically significant comparisons. Consider usefulness. Significant decreases are seen in columns 4 through 8 of Table 3. Time alone cannot cause this because time is controlled in columns 4, 6, and 8. Cohort alone cannot be the cause because the effect appears in columns 5 and 7, where cohort differences are eliminated. Practice cannot

Table 3
Summary of Main Effects From Analyses of 10 Personal Adjustment Variables

					Repeated measures			
					Two-level		Three-level	
	Cross-sequential		Time-sequential					
CAI variables	Time/ aging (1)	Cohort/ aging (2)	Time/ cohort (3)	Aging/ cohort (4)	Time/ aging/ practice (5)	Aging/ cohort (6)	Time/ aging/ practice (7)	Aging/ cohort (8)
Assessment of life	—	—	—	—	—	—	—	—
Satisfaction with accomplishments	Increased	—	—	—	Increased	—	—	—
Health	—	Decreased	—	Decreased	—	—	—	—
Friends	—	—	—	—	Decreased	—	—	—
Work	—	—	—	Decreased	Decreased	Decreased	Decreased	—
Economic security	—	Increased	—	Increased	—	Increased	—	—
Religion	—	—	—	—	—	Increased	—	—
Usefulness	—	—	—	Decreased	Decreased	Decreased	Decreased	Decreased
Happiness	—	—	—	—	Decreased	—	Decreased	—
Family	—	—	—	—	Decreased	—	—	—

Note. — indicates a nonsignificant effect. CAI = Chicago Attitude Inventory account for the effects in columns 4, 6, and 8. Aging might be the cause, however, because it is a factor in all of the comparisons in which sense of usefulness shows a significant effect.

By a similar logic, the effects for work must also be attributed to aging, but results for the other variables are not as clear. Time or aging could account for the effects on satisfaction with accomplishments; cohort or aging might explain the finding on health, economic security, and religion; and only cohort can be eliminated as a possible cause of changes in happiness, family, and friends.

It is logically possible that effects might be observed in which no single common element could be found. The first three columns might all show significant effects. In this case, it would be impossible to attribute the effect to a single cause, and the simplifying assumption would have to be dropped. Depending on the size and nature of the finding, an alternative would be to disregard (i.e., interpret as spurious) one or all of the three findings.

Internal consistency. Earlier it was pointed out that an ideal study design would incorporate several sources of data: longitudinal repeated measures, cross-sequential data from independent samples, and data from several cohorts measured simultaneously. These different sources of information can be used to estimate the size of different effects, provided of course that one is willing and able to make the necessary restrictive assumptions. In addition, however, these multiple sources of evidence can provide internal replications of findings. In the absence of theoretical guidance, results that emerge as the most internally consistent are those in which the most faith can be placed.

When an effect is replicated in an independent study, there is strong evidence that it is real and not due to chance or to some artifact. Internal consistency is a weaker criterion because the consistent results may be due to some shared artifact in the sample or data collection procedure. The criterion of internal consistency is thus best viewed as a basis for eliminating spurious findings. When we apply this standard in addition to logical elimination, we can see that only a few effects are consistently found. If we refer to Table 3, we can see that aging effects might be expected to appear in all columns except the third (time/cohort in the time-sequential design); that cohort effects should appear in columns 2, 3, 4, 6, and 8; that time-of-measurement effects should appear in columns 1, 3, 5, and 7; and that practice effects should be found in both columns 5 and 7. Usefulness and work show significant decreases in most of the comparisons in which aging is involved; economic security shows significant differences in three of five comparisons involving cohort; and both of the longitudinal comparisons, which involve practice effects, show effects for happiness. These results appear to be internally consistent, but most of the others do not. Logically, either aging or cohort could account for the significant effect seen on religion. A cohort effect for this variable, however, appears in only one of five possible comparisons and an aging effect in only one of seven. Under these circumstances, it seems best to view the one significant effect as due to chance.

External consistency. In interpreting effects, it is necessary to look beyond the data back to the constructs of interest. Conclusions about these constructs must be consistent with the data, but they should also fit in with other data and theories. In a narrow sense, the fact that findings support a prior hypothesis based on deductions from theory is the basic test of external consistency; in gerontological research, however, much of the work is exploratory, and theories are rarely sufficiently developed to support rigorous deductions. In these cases, researchers must consider the relation of their findings to other results in the literature and to recognized principles in the field.

Consider, for example, the conclusion that attitudes toward health do not decline with age. Is there any literature to support this? Indeed there is. Studies on age and self-reported health (Costa & McCrae, 1980; Markides & Martin, 1979) confirm this conclusion for most body systems. Without this corroborating evidence, the validity of the health attitude measure might have been called into question. On the other hand, we know of no literature demonstrating the practice effect on measures of happiness that we observe in these data. We could speculate that individuals wanted to give a favorable impression on their first visit and were more candid on the second, though this does not explain why similar drops were not seen in other variables. Part of scientific judgment is the ability to suspend judgment and to await further information before making any interpretation.

Rival hypotheses. Scientific reasoning is inductive and depends on the elimination of alternative explanations for observed effects. The elimination of cohort and time-of-measurement explanations for observed age differences or changes begins but certainly does not end this process. Researchers must entertain rival hypotheses and provide either data or arguments to support their choice among them. When an effect is attributed to "aging" or to "cohort" or to "time," it has not yet been explained. These are all "dummy" variables, representing some unidentified processes that they index and with which they covary. It is particularly dangerous to assume that an aging effect is equivalent to a maturational effect. The decline of positive attitudes toward work and usefulness is correlated with age, but it is unlikely to be the effect of aging itself. Declines in physical capacity might be the mediating variable in this relationship, but it is also possible that retirement is responsible. As long as society requires retirement at a certain age, a sense of uselessness may accompany age. If society changes the age of retirement, a corresponding change in sense of usefulness may occur. Any intervention depends on an understanding of the causal relationships among variables, and the separation of aging effects from generational differences and cultural changes is only the first step in gaining this understanding. Additional research is almost invariably required. Indeed, Maddox and Wiley (1976) argued that "The principal cause of the age-period-cohort *problem* is the habit of basing the assessment of age, period, and cohort effects on formally dependent operational definitions" (p. 21) and suggested that our understanding would be increased by a more direct examination of the environmental conditions that cohort and period are thought to index. They recommend that better judgment be exercised in the choice or operationalization of variables in order to minimize the judgment needed in the interpretation of results.

Perspective. Finally, one of the tasks of scientific judgment is in the evaluation of findings for their significance to an understanding of the field. This requires placing the results in some kind of perspective. Physicians normally ask about the "clinical significance" of findings in addition to their statistical significance. Do these results make any practical difference in the clinical condition or treatment of the individual? Incremental validity is a similar criterion, which asks whether a reported relationship adds any meaningful predictive value to the known set of factors that influence the variable. Perhaps the simplest method of assessing the importance of results is by calculating the proportion of the variance that is accounted for by a coefficient of determination or an omega squared. The decline in usefulness that we have spent so long attributing to an unknown but age-related cause amounts to an effect size of between .013 and .018 (Suskind & Howland, 1980), based on the two-time repeated measures analysis. Less than 2% of the variance in usefulness is accounted for by "aging." Critics might easily claim that we have belabored trivial findings.[3]

Nevertheless, perspective is more than effect size. Weighty theoretical distinctions might hinge on significant differences of a few percent, as the history of physics attests. The real question for judgment to address is, What is the import of these findings? Is an effect of such and such a size meaningful or important? Equally crucial is the question, What do the results as a whole suggest? Here it is often important to consider negative as well as positive results. From this perspective, the major implication of the present study is the finding that personal adjustment to aging, as measured by the CAI, is predominantly stable across age, time, and generations. Adjustment does not become noticeably more difficult with age, and that is a very noteworthy negative result. This finding is broadly consistent with the results of many other studies (e.g., Andrews & Withey, 1976; Campbell, 1976) and points to the necessity of investigating other variables, like personality, to obtain an understanding of personal adjustment. Ultimately science progresses because of judgments that certain topics are worth pursuing, and some deserve abandoning. Without a broader perspective in which to review results, these judgments cannot be made.

---

[3] See Cohen (1977) for other measures of effect size and Cooper (1981) for a discussion of "trivial" effects.

An Example of Clearer Effects

The effects of aging, period, cohort, and practice on scales of personal adjustment in the CAI are sufficiently complex that they require the explicit application of the principles of judgment. Nevertheless, lest the reader suppose that cross- and time-sequential analyses always yield such ambiguous results, a second example may be useful. Table 4 shows the results of similar analyses of age, period, and cohort effects on symptom complaints in the sections of the Cornell Medical Index (Costa & McCrae, 1980). If we disregard the unreplicated decrease in neurological complaints and assume as before that only one cause is operating, logical elimination shows that only aging could be responsible for the observed effects in this table. The conspicuous absence of any significant effect in the Time/Cohort column, where age is controlled, emphasizes this point. Further, the aging effects on Sensory, Cardiovascular, Genitourinary, and Total Complaints are internally consistent. Increases in Musculoskeletal complaints and Miscellaneous Diseases and decreases in poor Health Habits are not internally consistent and thus may be disregarded.

Again, however, all of these findings must be viewed in some perspective. An examination of effect size shows that the changes in symptom complaints associated with aging are quite small in magnitude, both absolutely and in comparison with effects of personality variables like neuroticism. On the other hand, it must be recalled that these

Table 4
Summary of Main Effects From Analyses of Cornell Medical Index (CMI) Physical Sections[a]

					Repeated Measures			
	Cross-sequential (N = 551)		Time-sequential (N = 637)		Two-level (N = 345)		Three-level (N = 118)	
CMI physical section	Time/ aging (1)	Cohort/ aging (2)	Time/ cohort (3)	Aging/ cohort (4)	Time/ aging/ practice (5)	Aging/ cohort (6)	Time/ aging/ practice (7)	Aging/ cohort (8)
Sensory systems	Increased	Increased	—	Increased	Increased	Increased	Increased	Increased
Respiratory	—	—	—	—	—	—	—	—
Cardiovascular	Increased	Increased	—	Increased	—	Increased	Increased	Increased
Digestive	—	—	—	—	—	—	—	—
Musculoskeletal	—	Increased	—	Increased	—	Increased	—	—
Skin	—	—	—	—	—	—	—	—
Neurological	—	—	—	Decreased	—	—	—	—
Genitourinary	Increased	Increased	—	Increased	Increased	Increased	Increased	Increased
Frequency of illness and fatigue	—	—	—	—	—	—	—	—
Miscellaneous diseases	Increased	—	—	Increased	—	Increased	—	—
Health habits	—	—	—	Decreased	—	Decreased	Decreased	—
Total physical complaints	Increased	Increased	—	Increased	—	Increased	Increased	—

Note. — indicates a nonsignificant effect.
[a] Data from Costa & McCrae (1980).

data were gathered from ambulatory volunteers, and attrition and self-selection must be credited with some role in diminishing the apparent role of age on self-reported health. Finally, the finding of small changes in predictable but select systems must be viewed in the context of pervasive stereotypes that depict aging individuals as subject to extensive and general deterioration. Data from a variety of analytic designs clearly show that this notion is untenable.

Conclusions

Psychologists, who have internalized the values of objectivity and mathematical rigor, normally welcome the appearance of statistical decision rules that promise to provide both. As an element in reaching conclusions, such rules are usually valuable and sometimes indispensable. The scientist, however, does not abdicate responsibility for good judgment by the mechanical application of any set of rules.

Even the ubiquitous determination of statistical significance requires judgment: Do the data meet the assumptions of the model? Is the alpha level sufficiently stringent? Is a one-tailed test really appropriate? A current controversy (Suskind & Howland, 1980) questions the sufficiency of statistical significance: Is effect size a better criterion of meaningful results?

When more elaborate decisions are required, the intelligent interpretation of statistical results is even more crucial. In discussing tests of invariance of models of measurement, Alwin and Jackson (1979) commented that "The acceptance of any of these models cannot be based on statistical grounds alone. . . . While statistical criteria provide one basis for making an interpretation of the data, the ultimate criteria for choosing a model depend on substantive considerations and objectives of research" (p. 103).

We have argued for a similar use of results from the application of the general developmental model. The quasi-experimental designs offered in that approach yield valuable data—if they are judiciously interpreted. Various writers have offered a number of different assumptive frameworks within which the data can be interpreted. The selection of an approach that is appropriate to the purpose of the research and the state of knowledge of the field is the first act of judgment required of the investigator who wishes to use these designs.

References

Adam, J. Sequential strategies and the separation of age, cohort, and time-of-measurement contributions to developmental data. *Psychological Bulletin*, 1978, *85*, 1309-1316.

Alwin, D. F., & Jackson, P. J. Measurement models for response errors in surveys: Issues and applications. In K. F. Schuessler (Ed.), *Sociological methodology 1980*. San Francisco, Calif.: Jossey-Bass, 1979.

Andrews, F. M., & Withey, S. B. *Social indicators of well-being: Americans' perceptions of life quality*. New York: Plenum Press, 1976.

Baltes, P. B. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, 1968, *11*, 145-171.

Baltes, P. B., Cornelius, S. W., & Nesselroade, J. R. Cohort effects in developmental psychology. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development*. New York: Academic Press, 1979.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. *Life-span developmental psychology: Introduction to research methods*. Belmont, Calif.: Wadsworth, 1977.

Birren, J. B., & Schaie, K. W. *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold, 1977.

Botwinick, J. *Aging and behavior*. New York: Springer, 1973.

Botwinick, J., & Arenberg, D. Disparate time spans in sequential studies of aging. *Experimental Aging Research*, 1976, *2*, 55-61.

Burgess, E. W., Cavan, R. S., & Havighurst, R. J. *Your activities and attitudes*. Chicago: Science Research, 1948.

Campbell, A. Subjective measures of well-being. *American Psychologist*, 1976, *31*, 117-124.

Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1963.

Cavan, R. S., Burgess, E. W., Havighurst, R. J., & Goldhammer, H. *Personal adjustment in old age*. Chicago: Science Research Associates, 1949.

Cohen, J. *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press, 1977.

Cooper, H. M. On the significance of effects and the effects of significance. *Journal of Personality and Social Psychology*, 1981, *41*, 1013-1018.

Costa, P. T., Jr., & McCrae, R. R. Somatic complaints in males as a function of age and neuroticism: A longitudinal analysis. *Journal of Behavioral Medicine*, 1980, *3*, 245-257.

Douglas, K., & Arenberg, D. Age changes, cohort differences, and cultural change on the Guilford-Zimmerman Temperament Survey. *Journal of Gerontology*, 1978, *33*, 737-747.

Gergen, K. J. Stability, change and chance in understanding human development. In N. Datan & H. W. Reese (Eds.), *Life-span developmental psychology: Dialectical perspectives on experimental research.* New York: Academic Press, 1977.

Glenn, N. D. Cohort analysts' futile quest: Statistical attempts to separate age, period and cohort effects. *American Sociological Review*, 1976, *41*, 900-904.

Glenn, N. D. Age, birth cohort, and drinking: An illustration of the hazards of inferring effects from cohort data. *Journal of Gerontology*, 1981, *36*, 362-369.

Havighurst, R. J. Validity of the Chicago attitude inventory as a measure of personal adjustment in old age. *Journal of Abnormal and Social Psychology*, 1951, *46*, 24-29.

Maddox, G. L., & Wiley, J. Scope, concepts and methods in the study of aging. In R. H. Binstock & E. Shanas (Eds.), *Handbood of aging and the social sciences.* New York: Van Nostrand Reinhold, 1976.

Markides, K. S., & Martin, H. W. Predicting self-rated health among the aged. *Research on Aging*, 1979, *1*, 97-112.

Mason, K. O., Winsborough, H. H., Mason, W. W., & Poole, W. K. Some methodological issues in cohort analysis of archival data. *American Sociological Review*, 1973, *38*, 242-258.

Rosenthal, R. Combining results of independent studies. *Psychological Bulletin*, 1978, *85*, 185-193.

Schaie, K. W. A general model for the study of developmental problems. *Psychological Bulletin*, 1965, *64*, 92-107.

Schaie, K. W. Quasi-experimental research design in the psychology of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging.* New York: Van Nostrand Reinhold, 1977.

Schaie, K. W., & Baltes, P. B. On sequential strategies in developmental research: Description or explanation? *Human Development*, 1975, *18*, 384-390.

Stone, J. L., & Norris, A. H. Activities and attitudes of participants in the Baltimore Longitudinal Study. *Journal of Gerontology*, 1966, *21*, 551-580.

Suskind, E. C., & Howland, E. W. Measuring effect magnitude in repeated measures ANOVA designs: Implications for gerontological research. *Journal of Gerontology*, 1980, *35*, 867-876.

Whitbourne, S. K., & Waterman, A. S. Psychosocial development during the adult years: Age and cohort comparisons. *Developmental Psychology*, 1979, *15*, 373-378.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

Butcher, J.N.; and Spielberger, C.D. (Eds.),
Advances in Personality Assessment, Vol. 4
Hillsdale, N.J.: Lawrence Erlbaum Assoc. (in press)

Concurrent Validation After 20 Years:
The Implications of Personality Stability for Its Assessment

Paul T. Costa, Jr.
and
Robert R. McCrae

Somewhere in the definition of "traits" given by most textbooks the phrase, "enduring dispositions" is likely to be found. These words mean that traits are to be distinguished from moods or other temporary states (Spielberger, 1972) that may affect a person. Traits are thought of as being characteristic, not of situations, seasons, or times-of-day, but of the individual at a particular point in his or her life.

But do these characteristics themselves endure, or do they change? That is not a matter of definition, but an empirical question. When gerontologists first began to examine the evidence on this question, their answers were mixed. A thorough and balanced review by one of the seminal writers in the field of aging and personality, Bernice Neugarten (1964, p. 188), came to the conclusion that "measures taken at long time interval tend to produce statistically reliable but relatively low correlations .... the implication is that there is at least as much change as stability."

Today we have the benefit of two decades of new evidence on this question, including the results of a number of major longitudinal studies published in the last six years. On the basis of this new data, a new conclusion is in order: in the course of normal aging, there is strong evidence of stability in the individual, and no consistent evidence at all of systematic change with age.

This conclusion is based not simply on more data, but on better quality data. In the early sixties, personality research was more often conducted with instruments of dubious validity and low reliability. Consider, for instance, the stability coefficients for the needs for achievement, affiliation, power, and aggression (r = .27, p = < .05) showed significant retest correlations in men, as did the need for achievement reported by Skolnick (1966). Over a twenty-year span from adolescence to adulthood, the needs for power (r = .34, p < .01, n = 44) and aggression (r = .24, p < .05, n = 49) in women. Other correlations did not reach significance. We see here the kind of "statistically reliable but relatively low correlations" that Neugarten based her conclusions on. Of course, higher stability is generally found _after_ the period of adolescence; but perhaps more importantly, TAT scores are notoriously unreliable. Correlations no higher than this are often seen in retests given weeks or months apart (Winter & Stewart, 1977). In fact, in view of the reliability of these measures, one might interpret Skolnick's data as strong evidence of stability.

More direct evidence, however, is now available from several studies. Leon, Gillum, Gillum, and Gouze (1979) examined the 30 year stability of MMPI scores for a sample of 71 men initially tested at age 40. Stability coefficients ranged from .28 for hypochondriasis and schizophrenia scales to .74 for social introversion, with a median correlation of .40. All correlations were statistically significant. The wide range of retest correlations in this data set might be interpreted to mean that some traits, like introversion, are stable, whereas others, like schizophrenia, are not. But an alternative explanation is to be found in the nature of the instrument and the sample. The MMPI is a measure of psychopathology, not normal personality. In a normal sample (like Leon's), the range on many scales is too restricted to show substantial correlations. Scales with more "normal" content, including Masculinity/femininity ($\underline{r}$ = .58) and Social Introversion ($\underline{r}$ = .74) show much higher correlations.

If this hypothesis is correct, we would expect uniformly high correlations when we examine normal subjects using a normal personality inventory such as the GZTS. Table 1 (Costa, McCrae, & Arenberg, 1980) shows data from the Baltimore Longitudinal Study of Aging, which has been conducted at the Gerontology Research Center for the past 25 years on a well-educated and generally healthy group of male volunteers. (Data collection on women began only in 1978.) Six-year retest correlations for men initially classified as young (17-44), middle (45-59) and old (60-85) are given in this table. Stability coefficients for such diverse traits as sociability, emotional stability, thoughtfulness and masculinity are all .7 to .8. Note also that what we might call the "stabilization" of personality occurs early in adulthood: the younger men are no less stable than the older.

By longitudinal standards, six years is not a very extensive interval. But much the same results were found when data from a second retest, twelve years later, were analyzed. Stability coefficients ranged from .59 to .87, and all were significant at the .001 level.

Strong as these data may be, they are not the upper limit of stability. Psychologists like Jack Block (1971) have argued that we should routinely correct for attenuation due to unreliability. Even though the GZTS is inherently more reliable than TAT measures, it too has an element of error, and a set of formulas developed by Heise (1969) allow us to estimate what the stability of "true" scores would be. Using these corrections, the estimated stability coefficients ranged from .80 to 1.00, with a median of .91 for the ten scales. Given perfect measures of the GZTS traits, these are the retest correlations we would expect to find after 12 years.

These findings have now been replicated in a number of other studies, including the Duke Longitudinal Study (Siegler, George, & Okun, 1979) and the Normative Aging Study in Boston (Costa & McCrae, 1977) where correlations as high as .84 were seen over a period of 10 years. In retrospect, we can argue that the conclusion of stability in adult personality was reached by another researcher thirty years ago, in a

Table 1

Six-Year Retest Coefficients for GZTS Scales in Different Age Groups

GZTS Scale	Total Group[a]	Young (17-44)[b]	Middle (45-59)[c]	Old (60-85)[d]
General activity	83	79	86	83
Restraint	71	68	69	70
Ascendance	82	79	84	80
Sociability	81	85	83	68
Emotional stability	74	84	72	77
Objectivity	71	69	73	70
Friendliness	77	78	78	75
Thoughtfulness	72	84	70	75
Personal relations	73	71	75	73
Masculinity	75	77	70	76

Note: Adapted from Costa, McCrae, & Arenberg, 1980. All correlations significant at $p < .001$; decimal points omitted. [a]N = 385-427. [b]N = 125-137. [c]N = 160-173. [d]N = 99-120.

somewhat different context. The best early longitudinal studies of objective test performance were conducted in the area of vocational interests by the great pioneer in that field, E. K. Strong. Given the close association of personality with vocational interests (Costa, Fozard, & McCrae, 1977; Holland, 1966), it is not surprising to hear that "Although permanence is less for young men than for older, it is still remarkably high for young men. The correlation of ....72 for 19-year-old college freshmen over an interval of 19 years is evidence of high permanency of interest" (Strong, 1951, p. 91). We can add that it is also evidence of the stability of personality.

Stability of Mean Levels

The retest correlations described so far are measures of one form of stability. They show that the relative ordering of individuals on a personality measure changes little from time to time. But they do not rule out the possibility that everyone is changing. This phenomenon is seen in the area of cognitive development among children, where marked growth in intellectual ability is seen from age 5 to age 15, but where relative individual differences are preserved: the bright 5-year-old is likely to be the bright 15-year-old. Physiological decline in functions like vital capacity would probably show a similar pattern of group change but high retest stability.

On the other hand, it is also possible that there are no systematic changes in personality traits, and that most individuals maintain the same absolute level of most traits.

We need only examine group means in order to determine if there are maturational increases or decreases in traits, and thus cross-sectional studies might give useful information on this point. A number of early cross-sectional studies had suggested that there might be developmental changes in personality, although the results were often inconsistent (Neugarten, 1977), and could generally be attributed to sampling biases or generational differences. Although they too have some interpretative problems, longitudinal studies provide a much sounder basis for discovering age-related changes. By measuring the same individuals repeatedly, we can use them as their own controls, and avoid the problems of subjects who represent different populations, or were socialized in different historical periods. Better yet are studies that longitudinally follow individuals initially of different ages. Both cross-sectional and longitudinal analyses can be conducted on these data, and we can feel relatively confident that an effect is due to aging when both cross-sectional age differences and longitudinal age changes are significant.

Probably the best study of changes in mean levels of personality traits was conducted by Douglas and Arenberg (1978) on the GZTS data for which retest correlations were presented in Table 1. In addition to cross-sectional and longitudinal analyses, Douglas and Arenberg also employed cross-sectional designs (Schaie, 1977), which provide yet another way to estimate age changes, independent of the possible "practice effects" introduced by longitudinal methods. Some of the scales showed generational differences or the effects of time of measurement, but only two of the ten GZTS scales--general activity and masculinity--showed patterns of results that could be interpreted as maturational change. And although statistically significant, the magnitude of these changes was so small (about one-eighth of one standard deviation over six years) as to be of no practical consequence.

These findings are typical of the results of longitudinal studies. In an analysis of the 16PF in the Normative Aging Study (Costa & McCrae, 1978), we found three scales with cross-sectional differences and two with longitudinal changes, but none that showed both. That meant that the cross-sectional differences were probably generational differences, and the longitudinal changes were probably due to practice effects, time of measurement differences, or simple error. Two-thirds of the scales showed neither age differences nor changes.

The 16PF was also used by Siegler, George and Okun (1979) in examining personality at the Duke Longitudinal study. Older generations scored lower on the 16PF measure of intelligence, probably because they are less well educated, and everyone improved on taking the intelligence test for a second time. Men became less guilt-prone and women more guilt-prone over the course of the study, but this sex-by-time interaction has never been replicated elsewhere and is probably due to chance. In the other 14 scales there is an unmistakable pattern: No age or cohort differences, no age or time or measurement changes. Leon et al. also noted little change in the mean level of MMPI scales over a thirty-year period.

These findings contradict some stereotypes of aging that depict old age as a time of ubiquitous decline. Hair goes, teeth go, memory goes--why not personality? Of course, there are age-associated conditions, such as dementia, that may have a profound impact on personality functioning. But in the absence of these conditions, personality does not ordinarily change, and in fact a marked change in personality may be a sign of organic pathology. As a general rule, old people do not succumb to depression or withdraw into introversion, or become increasingly conservative with age. Neither do they become more wise or mellow or emotionally integrated. some people at all ages show each of these characteristics, and those who show them in old age are most probably the same ones who showed them in their youth.

Alternative Interpretations

When we began to publish these findings and conclusions we were met with more than a little scepticism. Many researchers found our conclusions hard to believe, and suggested a number of alternative interpretations of the data that might account for the results. Inevitably, the methodology came under close scrutiny, and a number of objections were raised to the use of self-report instruments for the measure of personality. We therefore conducted a series of analyses designed to tell us if the stability we thought we saw was due to some artifact.

First we considered the problem of response sets. Individuals differ in the candor with which they answer personality questionnaires, and social desirability has been seen as an important determinant of test responses (Crowne & Marlowe, 1964; Edwards, 1957). Individuals also consistently differ in stylistic habits of responding regardless of item content: some people agree with almost any statement; some use extreme response categories, some prefer neutral answers. It has been proposed that it is these stylistic variables that are stable, not personality. The acquieser of today will be the acquieser of tomorrow, so of course he will receive the same scores.

Frankly, this is not a very plausible argument. different questionnaires typically involve different response formats as well as different items, and share in common only the substantive content. So it would be hard to explain why, for example, 16PF Extraversion predicts Eysenck's EPI Extraversion ten years later, as it does (Costa & McCrae, 1977). But data are better than arguments, so we set out to test the response set hypothesis.

In particular, we wanted to know if the high retest correlations we found over 6 years in the GZTS were inflated by the influence of stable response sets. To do this we recalculated the retest coefficients partialling out the variance due to response styles (Costa, McCrae, & Arenberg, in press). Although the GZTS itself has some built-in response tendency measures, we could not properly use them, since they are based on the same items as the trait scales. If an individual endorsed the same item on two occasions, we could not tell whether that consistency should be attributed to the stability of the trait or the response style measures from other instruments. From the Eysenck Personality Inventory (EPI; Eysenck & Eysenck, 1964) Lie scale we obtained a measure of social desirability; from the NEO Inventory (McCrae & Costa, in press), a personality questionnaire we have developed, we had measures of acquiesence and extreme responding. Under the hypothesis that stability of personality is merely stability of response sets, removing the variance attributable to social desirability, acquiesence, and extreme responding should substantially reduce retest coefficients. In fact, however, partialling out these stylistic variables did not reduce stability coefficients at all. It would appear that stable response sets cannot explain away stability in personality.

A second argument offered to explain consistent questionnaire responses is based on memory. Perhaps people want to appear consistent, recall how they responded the last time, and simply repeat their performance. Again, this argument is not convincing: why would so many people want so badly to appear consistent? And how could they remember specific item responses over so long an interval.

Again, an empirical study can be allowed the last word (Woodruff, in press). In 1969 Woodruff retested under two conditions subjects who had taken the California Test of Personality 25 years earlier. First they were asked to complete the questionnaire as it currently applied to them--the normal instructions for test administration. A few months later, however, they were again given the questionnaire, and this time they were asked to fill it out as they *recalled* filling it out in 1944. They were specifically instructed to use their memories to be as consistent as possible.

When the test was administered in the usual way, retest correlations were moderate for both men ($r = .58$, $p < .001$, $n = 53$) and women ($r = .58$, $p < .001$, $n = 24$). But when asked to rely on memory, subjects did much worse, with correlations of only .17 ($n = 30$, *ns*) for men, and .45 ($n = 18$, $p < .05$) for women. This is particularly intriguing finding. Not only does it eliminate the memory hypothesis from the interpretation of stability data--it also illustrates a profound fact about recollection. If you want to know what you were like 25 years ago, you are much more likely to be correct if you describe yourself as you are today than if you rely on your memory. Readers who are sceptical of the conclusion of stability because of recollections of change in their own lives should beware the distortions and exaggerations introduced by memory.

Finally, we considered one last artifact that could account for stability. People respond to personality questionnaires on the basis of the way they see themselves--their self-image or self-concept. But what if the self-concept is crystallized early in life? People may go on thinking of themselves as they were in young adulthood, when in fact they may have been changing gradually all along. Unlike the previous arguments, this one is plausible. Clinical case studies show that aspects of the self can change without a concomitant change in the self-concept, and the phenomenon of a crystallized self-concept has been noted by leading self-concept theorists (Rosenberg, 1979).

Further, this criticism, if sustained, would completely undermine the data collected for years by longitudinal researchers. No improvement in self-report measures, no experimental condition of administration can disprove this hypothesis. As long as we remain in the realm of self-reports, we are inescapably in the grip of the self-concept.

At this point research we had been conducting on spouse ratings of personality became useful. There is no theoretical reason to suppose that a husband's or wife's view of spouse's personality is crystallized; instead, they are likely to be responsive to the observable behavior of their spouse, and to any changes in it that have appeared over the years. Under this assumption, the correlation of spouse ratings with age should give an indication of age changes in personality completely independent of the self-concept.

But cross-sectional correlations of age with spouse-rated personality showed very small effects, none replicated across sexes. More elaborate analyses, reported at length elsewhere (McCrae & Costa, 1982b), suggest that even the small correlations observed are probably due to sampling bias or cohort effects rather than aging.

Clearly, a better basis for the conclusion would be longitudinal spouse-ratings; and in some respects, it would be better yet to seek outside opinions of experts at two different points in the life span. That is exactly what Jack Block (1971) has done in his Lives through time. A team of experts rated the adolescents in the Berkeley studies on the California Q-Sort; a different team rated them as 30-year-olds. The correlations of the two sorts, indicating the degree of agreement in the relative ordering of traits within the individual at two times, are almost all moderate to high in both men and women. Despite the fact that different raters are used, despite the fact that adolescents are compared with adults, the evidence for stability is clear. It is equally clear that the self-concept of the individual cannot account for these findings, since ratings are based on others' impressions, not one's own self-concept. Ratings and self-reports are in substantial agreement (McCrae, 1982); thus, we can return to self-report methods and measures with some assurance that they are not simply records of how personality was at the time of some hypothetical crystallization. Self-report methods, when used on volunteers who have little reason to distort their responses, appear to give an accurate account of personality as it is, and both self-reports and ratings concur in demonstrating considerable stability in a wide range of personality traits.

Validation Over Time

The most radical statement of the stability position would claim that the point at which personality is measured is irrelevant. Once the individual has reached full adulthood--say, by age 30--a single measurement of personality would suffice for a lifetime. Like eye color, date of birth, or years of education, personality might be considered a constant in the life of the individual.

There are, of course, a number of reasons to stop short of so sweeping a stand. It assumes a perfect correlations of tests over time, a condition rarely observed even when corrections for unreliability are made. It also assumes that there are no circumstances under which personality might regularly be expected to change. Evidence to date shows only that the events encountered over the course of a lifetime by volunteer subjects do not systematically produce change. But other circumstances, such as therapeutic interventions, cataclysmic events, or severe illnesses, might effect traits. The longitudinal study of personality in adulthood is simply too young a field to rule out these possibilities.

On the other hand, if we turn from the individual to the group, much stronger arguments can be made. On an aggregate level, there appear to be good grounds for claiming that the time at which measurements are made should not affect the results; and in particular, the relations between different tests ought not to depend on the times when they are administered. Data collected in the 1960's might be used to validate tests created in the 1970's, just as if administration had been contemporaneous.

Time interval does matter, however, when it is too short. Tests administered at the same sitting, or within a few days of each other, may share the kinds of distorting influences which lead to retest unreliability. Recent events may have temporarily elevated or depressed subjects' moods; the circumstances of the testing situation may induce anxiety or defensiveness; the close proximity of the tests may lead subjects to attempt to be consistent. Whenever two tests are given at the same time--the usual procedure when attempting to validate new measures--it can be presumed that correlations between the two will be inflated by these elements. In this case, some of the sources of unreliability artifactually increase, rather than decrease, the correlation between tests.

Because of these reasons, it could be offered as a general rule that studies of convergent validity should require as many differences in administration of test and criterion as possible: different experimenters, different locations, different times of day and seasons of administration should be sought. Under these conditions, we could be confident that the observed correlations were not spuriously inflated by transient and situationally-dependent moods.

In the rest of this chapter we will present data bearing on the construct validation of a new personality questionnaire, the NEO (Neuroticism-Extraversion-Openness) Inventory. The measures with which it will be correlated were collected by other investigators, under different conditions of administration--and they were administered from 2 to 21 years earlier. Because personality is stable, however, the logic will be that of a concurrent validational study.

An Example: The NEO Model

The studies to be reported here were made possible by the work of investigators who collected data over a period approaching a quarter century, and by the commitment of the subjects. Participants in the study are members of the Baltimore Longitudinal Study of Aging (BLSA), a community-dwelling, generally healthy group of volunteers who have agreed to return for testing at regular intervals. The sample has been recruited continuously since 1958, with most new subjects referred by friends or relatives already in the study. Until 1978, all participants were men, and only data from men are reported here. Subjects are well educated, with 93% high school graduates and 71% college graduates; nearly one-fourth have doctorate level degrees. Subjects range in age from 21 to 91.

The criteria of primary interest are from the Guilford-Zimmerman Temperament Survey (GZTS; Guilford, Zimmerman, & Guilford, 1976), administered to subjects from 1959 to 1979 on their first or second visit to the Gerontology Research Center. The GZTS was scored according to standard instructions, according to which individuals who used the "?" response for more than 3 of the 30 items in any scale were considered to have missing data for that scale (Douglas & Arenberg, 1978). Consequently, correlations with GZTS scales are based on a varying number of cases, although all cases for whom data were complete are used in the analyses. Although it was subsequently readministered, for purposes of simplicity, only first administration data is analysed here.

Over the same period of time, data were also collected on the Cornell Medical Index (CMI; Brodman, Erdmann, Lorge, & Wolff, 1949). Twelve sections deal with physical symptoms and six with psychiatric symptoms; these can be summed to score total physical and psychiatric complaints. Scores from the WAIS Vocabulary scale and the Army Alpha total (Costa & Shock, 1980) collected in the same two-decade period are also used.

All of these measures are to be used to consider the convergent and discriminant validity of three personality instruments administered in the past two years. Form A of the EPI, with scales for Extraversion, Neuroticism, and Lie, was mailed to subjects to be completed at home in September, 1979. The NEO Inventory, which will be the object of the major analyses, was mailed to subjects in February, 1980, and the NEO Rating Form was completed at home in August, 1980 by spouses of subjects whose husbands or wives were also participants in the BLSA. Because some spouses did not participate, rating data are available for only a subset of subjects.

The NEO Inventory measures six facets, or aspects, of three global domains of personality. Neuroticism is represented by anxiety, hostility, depression, self-consciousness, impulsiveness, and vulnerability. Extraversion includes scales for warmth, gregariousness, assertiveness, activity, excitement seeking, and positive emotions. Openness to Experience is measured in the areas of fantasy, aesthetics, feelings, actions, ideas, and values. Total scores for the three domains are obtained by summing the scores of the six facets in each. Data on reliability and factor structure are given elsewhere (McCrae & Costa, in press), and multi-method validation using spouse ratings has been reported (McCrae, 1982). The present analyses provide evidence of validity against different criteria; in addition, since the construct validity of the NEO scales is already fairly well established, they provide a demonstration of the feasibility of using data collected several years previously in validation studies.

Table 2

Correlations Between GZTS Scales Administered 1966-1979
And NEO Facets Administered in 1980

	GZTS Scales									
NEO Facets	G	R	A	S	E	O	F	T	P	M
Neuroticism:										
Anxiety	-14	-01	-25**	-25***	-67***	-55***	[illegible]***		-31***	-28***
Hostility	23**	-24**	07	-12	-47***	-44***	-53***	04	-26**	-24**
Depression	-04	-02	-38***	-38***	-58***	-52***	-26**	11	-29***	-26**
Self-Consciousness	-12	-04	-37***	-33***	-55***	-46***	-25**	02	-22**	-26**
Impulsiveness	12	-35***	16	11	-44***	-31***	-46***	-06	-27**	-13
Vulnerability	-31***	08	-44***	-34***	-39***	-38***	-03	01	-10	-26***
Extraversion:										
Warmth	11	-08	31***	56***	18*	15	08	10	12	-03
Gregariousness	31***	-25***	41***	52***	17	16	-05	-23**	16	00
Assertiveness	34***	-14	61***	41***	20*	15	-20	01	01	00
Activity	61***	-18*	20*	10	03	01	-16	01	-08	06
Excitement Seeking	41***	-34***	42***	21***	-13	-08	-33***	-05	01	02
Positive Emotions	32***	-28***	27**	30***	10	07	-06	10	-07	05
Openness:										
Fantasy	07	-12	10	-06	-30***	-20*	-22**	22**	-22**	-02
Aesthetics	19*	10	26**	26**	19*	15	13	31***	06	00
Feelings	22*	00	31***	21**	-09	-06	-24**	31***	-12	-15
Actions	22**	-10	31***	20*	03	05	00	09	08	24**
Ideas	12	13	29***	08	07	08	05	35***	09	11
Values	10	-12	23**	-05	-14	-13	-11	19*	-11	00

Note: GZTS scales are General activity (G), Restraint (R), Ascendance (A), Sociability (S), Emotional stability (E), Objectivity (O), Friendliness (F), Thoughtfulness (T), Personal Relations (P), and Masculinity (M). N = 138 to 152. Decimal points omitted. *$p < .05$; **$p < .01$; ***$p < .001$.

Correlations From Two Decades

Table 2 gives the correlations between GZTS scales administered between 1966 and 1979 and NEO facets administered in 1980, at least one year after the last GZTS. The average interval between the two tests was 9.1 years. Despite this considerable lapse of time, construct validity is very much in evidence. General activity from the GZTS is most strongly correlated with NEO activity; GZTS assertiveness with NEO assertiveness; GZTS sociability with NEO warmth; GZTS emotional stability and objectivity (negatively) with NEO anxiety; GZTS friendliness (negatively) with NEO hostility; GZTS thoughtfulness with NEO openness to ideas. At a more global level, the GZTS Extraversion scales (G, A, and S) are correlated chiefly with NEO Extraversion scales; GZTS Emotional Health scales (E, O, F, P, and M) show consistent negative correlations with NEO Neuroticism scales. Finally, the magnitude of the correlations requires comment: Almost all exceed .30, and several reach .60. Considering the somewhat different conceptions underlying the two instruments, these are large correlations.

But perhaps more impressive data is given in Table 3. Here the same correlations are presented for subjects who first completed the GZTS in the period from 1959 to 1966. At least 14, and as many as 21 years separate the two testing sessions; yet the correlations in Table 3 are almost identical to those in Table 2. Of the 86 significant correlations in Table 2, 67 (or 77%) are also significant in Table 3. Correlations for scales in the Neuroticism domain appear to be somewhat smaller in Table 3 than in Table 2, but those in the Extraversion domain are as large or larger. Indeed, a correlation of .66 is seen between GZTS activity and NEO activity in a sample of over 140 men after an average interval of 17.8 years.

In interpreting and comparing these two tables; it must be recalled that two different samples are involved. Even random samples are bound to differ in some respects, and these two groups of men are hardly random samples. There is some reason to believe that the initial group of volunteers differed systematically in intelligence and adjustment from later volunteers, and these differences might influence the correlations between the GZTS and NEO. In view of this, the similarities between the two tables are even more remarkable. Further analyses will combine subjects from the entire interval 1959-1979.

A joint factor analysis of the two instruments can demonstrate convergence at the domain level. The average interval between GZTS and NEO administrations was 13.5 years for the 235 men with complete data on both instruments. The GZTS has been shown to have three factors in a series of analyses (Guilford, Zimmerman, & Guilford, 1976); the same three factors were found in the present sample (McCrae, Costa, & Arenberg, 1980). We interpreted as neuroticism a factor Guilford labeled Emotional Health, and identified our own Extraversion with his Social Activity factor. The third factor in the GZTS, called Thinking Introversion,

Table 3

Correlations Between GZTS Scales Administered 1959-1966 And NEO Facets Administered in 1980

	GZTS Scales									
NEO Facets	G	R	A	S	E	O	F	T	P	M
Neuroticism:										
Anxiety	-07	05	-24**	-28***	-53***	-44***	-30***	11	-26**	-12
Hostility	08	-11	01	-10	-22**	-27***	-39***	-07	-24**	-10
Depression	-11	01	-20*	-25**	-37***	-44***	-28***	01	-32***	-18*
Self-Consciousness	-14	02	-40***	-36***	-37***	-41***	-21*	03	-31***	-22**
Impulsiveness	-01	-33***	-03	06	-25**	-21**	-27***	-10	-20*	-12
Vulnerability	-26**	-10	-39***	-35***	-17*	-18*	03	-20*	-16	-06
Extraversion:										
Warmth	24**	-09	29***	47***	04	10	01	16	09	-03
Gregariousness	28***	-26**	32***	57***	16	38***	-06	-15	20*	02
Assertiveness	47***	-02	56***	53***	25**	27***	-15	-01	13	12
Activity	66***	00	32***	30***	12	14	-16*	00	04	08
Excitement Seeking	27***	-33***	20*	37***	02	15	-24**	-17*	-02	06
Positive Emotions	33***	-19*	27***	42***	03	07	-11	07	-01	-07
Openness:										
Fantasy	-09	01	-09	-04	-28***	-23**	-04	17*	-09	00
Aesthetics	-02	21*	11	06	-07	-05	-04	22**	-02	-12
Feelings	27**	01	17*	23**	-22**	-15	-23**	38***	-09	-10
Actions	15	-14	19*	21*	18*	25**	10	-01	13	07
Ideas	-02	31***	08	01	-08	-12	08	35***	03	-04
Values	-03	-10	-02	05	00	07	01	-01	09	05

Note: GZTS scales are General activity (G), Restraint (R), Ascendance (A), Sociability (S), Emotional stability (E), Objectivity (O), Friendliness (F), Thoughtfulness (T), Personal Relations (P), and Masculinity (M). N = 140 to 152. Decimal points omitted. *$p < .05$; **$p < .01$; ***$p < .001$.

Table 4

Joint Factor Loadings for GZTS Scales Administered 1959-1979
And NEO Facets Administered 1980

	Factors		
	Neuroticism	Extraversion	Openness
GZTS Scales			
General activity		64	
Restraint		-50	35
Ascendance		67	
Sociability		67	
Emotional stability	-75		
Objectivity	-77		
Friendliness	-63	-38	
Thoughtfulness			67
Personal relations	-57		
Masculinity	-41		
NEO Facets			
Anxiety	77		
Hostility	67		
Depression	76		
Self-Consciousness	66	-35	
Impulsiveness	62		
Vulnerability	48	-47	
Warmth		48	
Gregariousness		67	
Assertiveness		71	
Activity		60	
Excitement seeking		64	
Positive emotions		59	
Fantasy	37		49
Aesthetics			67
Feelings		32	63
Actions			40
Ideas			72
Values			46

Note: Varimax-rotated principle components for 235 men. Decimal points omitted; loadings less than .30 not shown.

appeared to be the weakest of the three, since the two scales making it up, restraint and thoughtfulness, show only a very small simple correlation. We believe that restraint is more akin to low excitement seeking, a facet of Extraversion, and that thoughtfulness may be related to Openness.

These conceptual correspondences are fully corroborated empirically by the three-factor solution presented in Table 4. In fact, the emergence of Neuroticism, Extraversion, and Openness factors across two different instruments administered decades apart is a dramatic confirmation of two ideas we have been advancing for several years: that these are three major domains of personality, and that personality is stable.

Correlates of Three Domains

Table 5 presents a broader array of correlates in a somewhat simplified form. In place of the specific scales of the GZTS, second order domain scores are used. A Neuroticism score was formed by reflecting and summing emotional stability, objectivity, friendliness, and personal relation; and an Extraversion score was formed by summing general activity, ascendance, sociability and (reflected) restraint. The thoughtfulness scale is used to represent the third factor. Summary scores from the CMI and measures of intelligence and education are also included.

In addition to the NEO Inventory, Table 5 also presents correlations with the two other personality measures administered in the past two years: the EPI and the NEO Ratings. Overall domain scores have been used, and the columns have been organized by domain to facilitate comparisons across instruments.

Results are clear and in keeping with expectations about the nature of the constructs. Neuroticism is significantly related to (or, to preserve the time sequence, predicted by) GZTS Neuroticism, and CMI physical and psychiatric complaints. It is unrelated to measures of intelligence or education. Extraversion is strongly correlated with GZTS Extraversion, and there is some suggestion that introverts are more intelligent, at least in this sample. Extraversion is unrelated to CMI complaints. Openness to Experience, as measured by the NEO self-reports and spouse ratings, is clearly related to GZTS thoughtfulness, and shows a tendency to be associated with Extraversion; it is substantially independent of CMI complaints. Openness is also correlated with measures of intelligence and education (cf. Costa & McCrae, 1978). Note, however, that the correlations are not sufficiently large to permit the interpretation that Openness is nothing but intelligence.

For completeness, we have also included the Lie scale of the EPI, which shows small correlations with GZTS Neuroticism and CMI Psychiatric complaints. Although this might be interpreted as evidence that these

Table 5

Correlations of NEO, NEO Rating, and EPI Scores Administered in 1979-1980
With GZTS, CMI, and Intelligence Measures Administered 1959-1979

	Neuroticism			Extraversion			Openness		Lie
	NEO Self	NEO Rating	EPI	NEO Self	NEO Rating	EPI	NEO Self	NEO Rating	EPI
GZTS Factors[a]									
Neuroticism	61***	27**	66***	-01	-05	01	08	22*	-28***
Extraversion	-19**	03	-17	70***	59***	66***	19**	17	-10
Thoughtfulness	03	-04	17**	-01	-08	-14*	36***	26**	01
CMI Sections[b]									
Physical	34***	15	41***	-06	07	00	12*	14	-10
Psychiatric	41***	15	46***	-12*	-08	-05	05	05	-16**
Intelligence[c]									
WAIS Vocabulary	-02	-06	-01	-20***	-10	-29***	29***	28**	-08
Army Alpha Total	05	-07	04	-16**	-14	-25***	32***	27**	-21***
Education	-10	-12	-01	-01	20*	-11	24***	31***	-01

Note: [a]N = 258-300 for NEO Self-reports, 103-125 for Spouse ratings, 288-330 for EPI. [b]N = 303 for NEO Self-reports, 123 for Spouse ratings, 332 for EPI. [c]N = 261-316 for NEO Self-reports, 104-130 for Spouse ratings, 286-351 for EPI. Decimal points omitted. *p < .05; **p < .01; ***p < .001.

Table 6

Correlations of CMI Psychiatric Sections Administered 1959-1979
With NEO Neuroticism Facets Administered 1980

	Cornell Medical Index "Moods and Feelings"					
	Inadequacy	Depression	Anxiety	Anger	Sensitivity	Tension
NEO Neuroticism						
Anxiety	25***	27***	27***	35***	34***	30***
Hostility	04	24***	21***	28***	37***	15*
Depression	20***	32***	15**	32***	27***	24***
Self-Consciousness	27***	32***	11	29***	24***	25***
Impulsiveness	10	10	11	14*	14*	10
Vulnerability	34***	29***	12*	25***	19***	31***

Note: Decimal points omitted. N = 303. *$p < .05$; **$p < .01$; ***$p < .001$.

measures are somewhat distorted by the influence of social desirability, other analyses suggest that the Lie scale is instead substantively related to Neuroticism (McCrae & Costa, in press).

Correlations with NEO ratings are generally smaller than those with self-reports, and occasionally fail to reach statistical significance, in part because they are based oin considerably fewer cases. Considering that the correlations shown here cross instruments, methods of measurement, and 2 to 20 years, however, the pattern of results should be convincing. The correlation of .59 between GZTS Extraversion and spouse-rated NEO Extraversion is particularly remarkable.

Finally, Table 6 gives correlations between individual CMI psychiatric section scores and NEO Neuroticism facets. Since psychiatric symptoms are the extreme expression of normal Neuroticism, we would expect most of the correlations to be positive. Examination of the individual sections, however, gives an opportunity to check the discriminant validity of the NEO Neuroticism facets.

The CMI is a less-than-ideal instrument for assessing psychiatric symptoms, and a number of superior measures, like the SCL-90 (Derogatis, 1977), now exist. In 1959, however, when data collection began at the BLSA, it was not a bad choice. As a criterion against which to validate the NEO scales, its chief drawback is that many of the items reflect such sever pathology that endorsement frequencies are very low. Out of a possible total of 45 items, the average BLSA participant endorsed only 2.23. In addition, some of the descriptive labels given to the scales are somewhat misleading: Sensitivity refers to "touchiness" or irritability, and tension involves severe signs of anxiety including nightmares and breaking out in a cold sweat.

In view of these considerations, the evidence of discriminant validity within the domain of Neuroticism is fairly strong. CMI inadequacy and tension are most strongly related to NEO vulnerability; CMI depression to NEO depression and self-consciousness; CMI anxiety to NEO anxiety; and CMI sensitivity to NEO hostility. The only unexpected finding is with CMI anger, which is more highly correlated with NEO anxiety, depression, and self-consciousness than with NEO hostility.

Some Implications and Issues

Historical Effects on Test Data

As the value of longitudinal studies becomes more apparent, we can expect more researchers to collect and archive personality data, or to retrieve and retest subject populations measured some time before (e.g., Woodruff, in press). Longitudinal methodologists have pointed out that there are potential problems in the use of such data, since the meaning of the test may have changed in the intervening years. This phenomenon can readily be envisioned in the case of attitudinal research: attitudes toward women that were radical in the 1950's might be middle-of-the-road today.

In the case of personality measures, this concern appears to be less well-founded. Over the past 20 years, scales measuring the major dimensions of personality do not appear to have altered in meaning substantially. Subjects today appear to respond to test items in much the same way as they did in the 1950's and 1960's. Basic aspects of temperament are relatively impervious to historical changes, and longitudinal research can use older personality data with some confidence that the meaning of test items has not changed.

Indeed, it seems more likely that what may have changed is our *conceptualization* of what older tests measure. For example, the Bernreuter (1935) inventory contained a scale labelled "introversion" that we might now conceptualize instead as social anxiety, a part of Neuroticism. Such hypotheses are testable if we readminister older instruments and revalidate them against better conceptualized and psychometrically more sophisticated measures. Since the relations between tests appear to change little over an interval of years, we could use the results of contemporary studies to reinterpret older findings on the basis of our current understanding of the nature of the constructs they measure.

Accumulating Data on a Single Sample

When we began to work with longitudinal studies and their pools of indefinitely-retestable subjects, we soon realized that one of their major strengths had usually been overlooked by the advocates of longitudinal research. Once a battery of measures has been collected on a sample, the addition of each new measure permits an entire study on the relations of that measure to all the others. The hundred-and-first scale may take only a few minutes to administer, but it will provide information on 100 different convergent or discriminant relations. Instead of gathering a new sample whenever one wishes to investigate the relation between two tests, administering new tests to the same subjects previously tested becomes increasingly profitable with each new measure.

Some researchers have recognized this fact and attempted to capitalize on it by undertaking massive testiq programs in a few days or weeks (e.g., Coan, 1974). In addition to being extremely burdensome to subjects, this strategy suffers from the possible confounding effects of temporary states or situational influences that all the tests will share in common. It is both more humane and more informative to stretch the period of data gathering out over a period of months or years.

The data in this chapter suggest that such a strategy might well be carried out over decades. Personality researchers might select a group of subjects and test them every six months for 50 years, using whatever personality measures and constructs seemed most appropriate at the time. The result would be an immense archive of information, providing definitive studies on the relations between traits and their successive operationalizations. In addition, of course, it would be a longitudinal record, allowing studies of antecedents and outcomes that would be invaluable in showing the influence of personality on the shaping of the life course (McCrae & Costa, 1982a).

Of course, the accumulation of so much data on a single sample would present certain statistical problems. The probability of spurious correlations would be high, and researchers would need to adopt stringent probability requirements and rely heavily on internal replication. In addition, the sample itself would have to be very large--probably several thousand to begin with--and ought certainly to be more representative of the population in general than most longitudinal studies have been. But these kinds of populations do exist (e.g., Bachman, O'Malley, & Johnston, 1978), and systematic study of them would be well worth the effort involved in recruiting and maintaining the sample.

Stability in the Individual

Finally, the stability of personality is sufficiently high to warrant judicious use of personality data as a part of individuals' permanent records. Physicians, for example, might want to include personality scores taken in young adulthood as part of the standard medical history. With some training in psychology, this information might be of use to the physician in dealing with his or her patient--for example, in interpreting the significance of medical complaints (Costa, Fleg, McCrae, & Lakatta, 1982; Costa & McCrae, 1980b). In addition, documented changes in personality might provide useful diagnostic information in some cases. Certainly the research applications are attractive: With such records on most of the population, we could resolve once and for all many questions about the medical consequences of such dispositions as anxiety, depression, impulsiveness, or activity level.

Although these kinds of records are frequently kept in Europe, Americans have traditionally been wary of official dossiers on them, and the potential for abuse is significant: an individual who scored high on a measure of neuroticism once, at age 25, might be labeled "neurotic" for a lifetime. Trained psychologists have to be reminded that all scores are subject to error of measurement, and that Neuroticism is a normal personality dimension; laymen could easily misunderstand this information. Even the simple statement that personality is stable can be dangerous if it confirms people in mistaken preconceptions or deprives individuals of hope for personal growth and fulfillment. These ethical issues will become increasingly salient as the evidence of stability in personality accumulates.

Summary

A number of previous articles have provided evidence on the stability of personality traits, and a few (e.g., Costa & McCrae, 1980a) have used longitudinal data to show that personality variables preceeded, and thus may have caused, later outcomes. As far as we know, however, this article presents the first systematic attempt to assess the convergent and discriminant validity of a personality inventory using criteria collected as much as 20 years before. The results show strong patterns of construct validity for the NEO Inventory and NEO Rating Form, expecially since the correlations cannot be due to transient states or situational influences shared in common by the measure and the criteria. Indirectly, the results also reaffirm the conclusion that personality is extremely stable in adulthood.

Acknowledgements

Portions of this paper were presented by the senior author at a seminar given to the Department of Psychiatry, Johns Hopkins University School of Medicine, on November 1, 1982. We gratefully acknowledge the contributions of investigators and technicians in the Baltimore Longitudinal Study of Aging, particularly Dr. David Arenberg, for the collection of much of the data reported here.

References

Bachman, J. G., O'Malley, P. M., & Johnston, J. *Adolescence to adulthood: Change and stability in the lives of young men*. Ann Arbor, MI.: Institute for Social Research, 1978.

Bernreuter, R. G. The theory and construction of the Personality Inventory. *Journal of Social Psychology*, 1933, *4*, 387-405.

Block, J. *Lives through time*. Berkeley, CA.: Bancroft Books, 1971.

Brodman, K., Erdmann, A. J., Jr., Lorge, I., & Wolff, H. G. The Cornell Medical Index: An adjunct to medical interview. *Journal of the American Medical Association*, 1949, *140*, 530-534.

Coan, R. W. *The optimal personality*. New York: Columbia University Press, 1974.

Costa, P. T., Jr., & McCrae, R. R. Age differences in personality structure revisited: Studies in validity, stability, and change. *Aging and Human Development*, 1977, *8*, 261-275.

Costa, P. T., Jr., & McCrae, R. R. Objective personality assessment. In M. Storandt, I. C. Siegler, & M. F. Elias (Eds.), *The clinical psychology of aging*. New York: Plenum Press, 1978.

Costa, P. T., Jr., & McCrae, R. R. The influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, 1980, *38*, 668-678. (a)

Costa, P. T., Jr., & McCrae, R. R. Somatic complaints in males as a function of age and neuroticism: A longitudinal analysis. *Journal of Behavioral Medicine*, 1980, *3*, 245-257. (b)

Costa, P. T., Jr., & Shock, N. W. New longitudinal data on the question of whether hypertension influences intellectual performance. In M. F. Elias & D. H. P. Streeten (Eds.), *Hypertension and cognitive processes*. Mt. Desert, ME: Beech-Hill Publishers, 1980.

Costa, P. T., Jr., Fleg, J. L., McCrae, R. R., & Lakatta, E. G. Neuroticism, coronary artery disease and chest pain complaints: Cross-sectional and longitudinal studies. *Experimental Aging Research*, 1982, *a*, 37-44.

Costa, P. T., Jr., Fozard, J. L., & McCrae, R. R. Personological interpretation of factors from the Strong Vocational Interest Blank scales. *Journal of Vocational Behavior*, 1977, *10*, 231-243.

Costa, P. T., Jr., McCrae, R. R., & Arenberg, D. Enduring dispositions in adult males. *Journal of Personality and Social Psychology*, 1980, *38*, 793-800.

Costa, P.T., Jr., McCrae, R. R., & Arenberg, D. Recent research on personality and aging. In K. W. Schaie (Ed.), *Longitudinal studies of adult development*, New York: Guilford Press, in press.

Crowne, D., & Marlowe, D. *The approval motive*. New York: Wiley, 1964.

Derogatis, L. R. *SCL-90 Manual-I*. Baltimore: Leonard R. Derogatis, 1977.

Douglas, K., & Arenberg, D. Age changes, cohort differences, and cultural change on the Guilford-Zimmerman Temperament Survey. Journal of Gerontology, 1978, *33*, 737-747.

Edwards, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

Eysenck, H. J., & Eysenck, S. B. G. *Manual of the Eysenck Personality Inventory*. London: University Press, 1964.

Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. *The Guilford-Zimmerman Temperament Survey Handbook: Twenty-five years of research and application*. San Diego, CA.: EdITS Publishers, 1976.

Heise, D. R. Separating reliability and stability in test-retest correlation. *American Sociological Review*, 1969, *34*, 93-101.

Holland, J. L. *The psychology of vocational choice: A theory of personality types and model environments*. Waltham, MA: Blaisdell, 1966.

Leon, R. R., Gillum, B., Gillum, R., & Gouze, M. Personality stability and change over a 30 year period--middle age to old age. *Journal of Consulting and Clinical Psychology*, 1979, *23*, 245-259.

McCrae, R. R. Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology*, 1982, *43*, 293-303.

McCrae, R. R., & Costa, P. T., Jr. Aging, the life course, and models of personality. In T. Field (Ed.), *Review of Human development*, New York: Wiley, 1982. (a)

McCrae, R. R., & Costa, P. T., Jr. The self-concept and the stability of personality: Cross-sectional comparisons of self-reports and ratings. *Journal of Personality and Social Psychology*, 1982, *43*, 1282-1292. (b)

McCrae, R. R., & Costa, P. T., Jr. Joint factors in self-reports and ratings: Neuroticism, extraversion, and openness to experience. *Personality and Individual Differences*, in press.

McCrae, R. R., & Costa, P. T., Jr. Social desirability and neuroticism: Which contaminates which? In preparation.

McCrae, R. R., Costa, P. T., Jr., & Arenberg, D. Constancy of adult personality structure in adult males: Longitudinal, cross-sectional and times of measurement analyses. *Journal of Gerontology*, 1980, *35*, 877-883.

Neugarten, B. L. Personality change over the adult years. In J. E. Birren (Ed.), *Relations of development and aging*. Springfield, IL: Charles C. Thomas, 1964.

Neugarten, B. L. Personality and aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. Reinhold, 1977.

Rosenberg, M. *Conceiving the self*. New York: Basic Books, 1979.

Schaie, K. W. Quasi-experimental research designs in the psychology of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold, 1977.

Siegler, I. C., George, L. K., & Okun, M. A. Cross-sequential analysis of adult personality. *Developmental Psychology*, 1979, *15*, 350-351.

Skolnick, A. Stability and interrelationships of thematic test imagery over twenty years. *Child Development*, 1966, *37*, 389-396.

Spielberger, C. D. Anxiety as an emtional state. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research*, Vol. I. New York: Academic Press, 1972.

Strong, E. K., Jr. Permanence of interest scores over 22 years. *Journal of Applied Psychology*, 1951, *35*, 89-91.

Winter, D. G., & Stewart, A. Power motive reliability as a function of retest instructions. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 436-440.

Woodruff, D. Age and personality: A twenty-five year follow-up. *Experimental Aging Research*, in press.

Malatesta, C.; and Izard, C. (Eds.),
Affective Processes in Adult Development and Aging.
Beverly Hills: Sage Publications (in press)

Personality as a Lifelong Determinant of Well-Being

Paul T. Costa, Jr.
and
Robert R. McCrae

Despite the folk wisdom implicit in such sayings as "money can't buy happiness", one of the most difficult lessons for students of well-being to learn has been the relative independence of objective and subjective well-being (Lawton, this volume). Most of us firmly believe that we would be happier if we had better health, greater prestige, lower taxes, or more security. The life stories of a hundred movie stars or lottery winners notwithstanding, we feel certain that we would appreciate our new-found blessings and live happily ever after if only we got whatever it is we currently want.

Conversely, we expect we would be less happy if we had poorer health, lower status, less money, fewer friends. And since all these--together with decline in sexual vigor, loss of physical beauty, and the imminence of death--are associated with age, most of us regard the approach of old age with attitudes ranging from resignation to gloom to desperation. Surely old people must be among the most unhappy of all groups!

Yet survey research has shown again and again that the elderly are not particularly unhappy--in fact, may be more satisfied with life than are younger men and women (Andrews & Withey, 1976; Campbell, Converse, & Rogers, 1976; Herzog, Rodgers, & Woodworth, 1982). Should that be taken as evidence that life is better for the aged than we imagined, that there are compensatory gains? Does it suggest that their responses to measures of subjective well-being are meaningless, mere rationalizations, a brave but false front intended to deny the inevitable? Or is happiness perhaps determined by other factors that do not change with age?

These questions are of more than academic interest, but even as intellectual problems they are fascinating. Their answers depend on an understanding of the relations between environment and cognition, cognition and affect, affect and personality. None of these is fully understood, but in recent years considerable progress has been made. This chapter will review some of the evidence and present results of new studies confirming, extending, and in some cases modifying a model of personality and subjective well-being we have been developing over the past few years. Our new data come from the Augmented Baltimore Longitudinal Study of Aging (ABLSA; McCrae, 1982), a well-educated, generally healthy, community-dwelling group of volunteers who cover the full adult age range and have been studied for as much as 25 years. Other research, done here (Costa, McCrae, & Norris, 1980; Lawton, this volume) and abroad (Warr, Barter, & Brownbridge, 1983) supports the generalizability of the basic model.

Well-being Social Desirability and Adaptation

Before we attempt to explain the paradox of happiness in old age, we should be sure that the reports that create it are trustworthy. When self-report results contradict experimenter expectations, social desirability or defensive responding is often urged as an explanation of the discrepancy. Tamir (1982), for example, thinks that her failure to find much evidence of a mid-life crisis can be accounted for by socially desirable responding in middle aged men, and Herzog, Rodgers, and Woodworth (1982) raise this as a possible explanation for the high reported life-satisfaction of the elderly. Because most research on well-being is done using self-reports (and must be if subjective well-being is the variable of interest), this charge would be serious if supported.

But there are at least three good reasons for doubting that it is true. First, despite widespread beliefs, social desirability as a response style has rarely proven to be a significant threat to the validity of self-report measures (Dicken, 1963; McCrae & Costa, in press). Second, the hypothetical distortion we attribute to the elderly would also have to be attributed to minorities, the poor, the undereducated, and many other groups who might be supposed to be lower in well-being, but who show relatively small differences in most studies (e.g., Andrews & Withey, 1976). When so many groups that should be unhappy are not, some process more fundamental than mere responding bias is suggested. Third, there is an alternative explanation that plausibly accounts for the general lack of age differences: adaptation theory.

Adaptation level (Helson, 1964) is a general theory of perception that holds that an individual's estimation of quantity depends on his or her recent experience rather than on absolute standards. A brass band may sound "loud" to a devotee of chamber music, but not to a rock musician. The same model has been extended to more abstract perceptions, including satisfaction with housing, perceived health, and general well-being (Brickman & Campbell, 1971). According to this model, changes in the quality of life lead to increases in happiness only temporarily, until the neutral point of comparison has been reset. Brickman and Campbell refer to this phenomenon as the "hedonic treadmill", and they imply that the rising standard of living that America enjoyed in the fifties and sixties could not be expected to produce a concomitant rise in well-being. The good news, of course, is that the declining standard of living of recent years has brought less misery than it would have had we not all adjusted our expectations downward.

Brickman, Coates, and Janoff-Bulman (1978) gave a dramatic example of the power of adaptation by comparing lottery winners with paraplegics and showing little difference in well-being. A more persuasive test of adaption level theory would utilize a longitudinal design: we would predict large initial differences between groups who had recently experienced tragic or fortunate events, followed by gradual readjustment to a neutral point. In the absence of such studies, adaptation level theory remains an attractive and intuitive hypothesis.

Note, however, that this theory calls for a reinterpretation of most measures of psychological well-being. Well-being has been measured as satisfaction in different areas of life, as a balance of positive to negative affects (Bradburn, 1969), or as a simple avowal of happiness (Gurin, Veroff & Feld, 1960). But all these indices of subjective well-being have customarily been interpreted as reflections of the objective quality of life. Happy, satisfied people must be leading rewarding lives--or so it was assumed. But if we take adaptation level theory seriously, we will interpret these signs differently. We will say instead that, compared to the recently established neutral point, happy and satisfied people are better off than they were. Happiness becomes entirely relative, and the basis of comparison is recent past experience. We could not, under this theory, tell whether the quality of life was good or bad in any absolute sense; we could only infer that it had recently improved or deteriorated. As long as negative age-changes are sufficiently gradual, they might have no noticeable impact on subjective well-being.

Personality and Well-Being

Although there is doubtless a good deal of truth in this model of happiness, there are also some serious flaws. Studies of the stability of happiness show that something else must be going on as well.

If well-being were nothing but temporary perturbations around a neutral point, then we would expect most individuals to be unhappy as often as happy. In fact, retest correlations of well-being measures ought to be consistently negative, because the happier one is today, the higher the neural point becomes, and the more likely one is to be unhappy tomorrow. Happiness could be sustained only for the fortunate few whose lives every day in every way were getting better and better.

Yet data from a number of studies clearly show that, while not immutable, happiness is relatively stable in individuals. We recently readministered a battery of well-being measures to our ABLSA subjects after an interval of two years. The battery included the Bradburn (1969) Affect Balance Scales, that yield separate scores for Positive Affect (PAS), Negative Affect (NAS) and the difference of these two, Affect Balance; a Satisfaction index that asks whether subjects are not at all, somewhat, or very satisfied with fourteen areas of living (housing, city, government, work, leisure, appearance, sex, health, marriage, family, finances, friends, self-respect, and faith); and the Delighted-Terrible (D-T) scale of Andrews and Withey (1976), a single item which asks subjects to evaluate their life as a whole. Retest correlations for these five measures range from .47 to .63 (&N = 473, $p < .001$). Similar values have been reported elsewhere for different intervals (Costa & McCrae, 1981; Palmore & Kivett, 1977). Well-being indicators may well be sensitive to short-term changes in life quality, but they are also influenced by more enduring conditions.

As Bradburn and others have consistently demonstrated, positive affect and negative affect are independent contributors to global well-being. In part this appears to be because the external events that elicit them are independent (Warr, Barter, & Brownbridge, 1983), like the hygenic and intrinsic factors that have been identified as contributors to job satisfaction (Herzberg, Mausner, & Snyderman, 1959). But another reason for their independence is that positive affect and negative affect are differentially related to the two personality dimensions of extraversion and neuroticism. We believe the influence of these personality dimensions contributes to stability in well-being.

In data from Boston's Normative Aging Study (Costa & McCrae, 1980a), we showed that Positive Affect Scale (PAS) scores in a large sample of adult men were consistently related to personality traits which together formed the broad domain of extraversion. Negative Affect Scale (NAS) scores were unrelated to extraversion, but were predicted by traits in the domain of neuroticism. Affect Balance scores, along with other global measures of well-being (hopelessness, personal security, and a life satisfaction index) were related positively to extraversion and negatively to neuroticism.

Well-Being and Personality in Adult Men and Women: New Data

In our more recent research we have employed a new instrument--the Neuroticism-Extraversion-Openness or NEO Inventory--to assess personality (Costa & McCrae, 1980c; McCrae & Costa, in press). The NEO Inventory provides scores on six different facets or aspects for each of three distinct, global domains of personality. Domain scores for N, E, and O are formed by summing the respective six subscales. The inclusion of separate facets allows us to determine which specific aspects of neuroticism and extraversion are responsible for the associations with well-being.

The right panel of Table 1 shows the relations between NEO scales and well-being measures administered in 1979 to a sample of 350 men aged 25 to 91. As hypothesized, PAS is related chiefly to traits in the domain of extraversion, NAS to traits in the domain of neuroticism. Both extraversion and neuroticism are associated with overall well-being, whether measured as affect balance, satisfaction, or the single-item D-T scale.

An examination of specific facets shows that all of the neuroticism facets or traits, especially anxiety and depression, are related to well-being. Individuals who are chronically anxious, hostile, depressed, self-conscious, impulse-ridden, and vulnerable to stress are (not

Table 1

Correlations Between Self-Reported NEO Scales and 1979 Well-Being Measures

NEO Scale	Men[a] PAS	NAS	Balance	Satis-faction	D-T Scale	Women[b] PAS	NAS	Balance	Satis-faction	D-T Scale
Anxiety	-14**	40***	-37***	-35***	-36***	-18**	34***	-35***	-27***	-29***
Hostility	-09	32***	-27***	-28***	-24***	-13*	34***	-32***	-29***	-31***
Depression	-20***	44***	-44***	-43***	-42***	-28***	59***	-59***	-41***	-44***
Self-Consciousness	-10	32***	-29***	-27***	-27***	-16*	38***	-36***	-31***	-30***
Impulsiveness	06	29***	-15**	-23***	-12*	-08	39***	-32***	-30***	-24***
Vulnerability	-23***	20***	-29***	-31***	-30***	-19**	43***	-42***	-34***	-40***
Neuroticism	-15**	45***	-41***	-42***	-38***	-23***	54***	-52***	-42***	-43***
Warmth	23***	-11*	23***	22***	25***	17**	-10	17***	24***	23***
Gregariousness	14*	00	09	06	11*	05	00	03	10	05
Assertiveness	24***	-13*	24***	17**	17**	19***	-11	19**	19**	16**
Activity	30***	00	20***	10	04	16**	13*	01	08	06
Excitement Seeking	14**	13*	01	-05	01	05	20**	-11	-09	-07
Positive Emotions	40***	-01	28***	22***	35***	39***	-12*	33***	25***	35***
Extraversion	36***	-02	26***	17**	22***	28***	01	17***	20**	21***
Fantasy	06	23***	-12*	-15**	-06	-04	23***	-19**	-14*	-15*
Aesthetics	20***	05	10	09	08	25***	00	15*	01	06
Feelings	24***	18***	04	08	09	08	21***	-10	-04	00
Actions	13*	02	08	07	13*	17**	-01	12	12	04
Ideas	14*	02	08	06	04	09	-06	10	04	01
Values	04	18***	-09	-04	-09	03	04	-01	-09	-09
Openness	21***	18***	02	03	05	15*	10	02	-03	-03

Note: [a]N = 350 for Bradburn scales; 344 for Satisfaction; 342 for D-T Scale. [b]N = 256 for Bradburn scales; 256 for Satisfaction; 250 for D-T Scale. Decimal points omitted. *$p < .05$; **$p < .01$; ***$p < .001$.

surprisingly) likely to have an unfavorable affect balance, and to be dissatisfied with life. In the case of extraversion, the relationships seem more differentiated. Interpersonal warmth leads to greater happiness, but mere gregariousness often does not, despite the fact that these two aspects of sociability are highly correlated. Assertiveness and the predisposition to experience positive emotions consistently correlate with happiness and life satisfaction, but excitement seeking (which is akin to Zuckerman's (1979) sensation seeking) does not. Activity is related to positive affect, but not to life satisfaction.

As in all our previous studies on this topic, the data in the right panel of Table 1 derive only from men. Can the same model be applied to women? The left hand panel of Table 1 suggests that it can. Personality data from 256 women aged 24 to 96 are presented here, and the same general pattern is found. In this sample, the relation between NAS and neuroticism is exceptionally high ($\underline{r} = .54$); correlations with extraversion are comparable to those found in men. As in men, depression is the facet contributing most to the prediction of NAS; warmth, assertiveness, and positive emotions are most strongly related to PAS.

The data from both men and women call for a reinterpretation of the relations between extraversion and well-being. In many respects extraversion appears to be a more complex dimension than neuroticism, compounding the conceptual difficulties that arise from its frequent confusion with the Jungian concept of the same name (McCrae, 1983). Eysenck and Eysenck (1968) argued that sociability and impulsivity were identifiable components of a larger extraversion factor, and factor analyses of the NEO Inventory (McCrae & Costa, in press) found dominance, affiliation, and impulsivity components. Excitement seeking and gregariousness, elements placed in the impulsivity component, do not appear to be related to subjective well-being. This suggests that the arousal associated with thrill-seeking and the social stimulation of crowds does not contribute to well-being. Instead, friendliness, self-confidence, and cheerfulness appear to be the key facets of extraversion responsible for the association. (Incidentally, these distinctions show the potential utility of the multifaceted approach to measuring personality domains embodied in the NEO Inventory.)

Wessman and Ricks (1966; Wessman, 1977) point out that hedonic level--the average level of day to day mood--is independent of mood variability. They found that individuals who experienced the most dramatic mood shifts were neither more nor less happy than stable individuals. They were, however, characterized by such personality traits as imagination, enthusiasm, and openness, whereas less moody individuals were rigid, cautious, and closed. Openness to experience seems to have similar effects on subjective well-being. Individuals who are open to experience seek variety and novelty, and have an appreciation for the intrinsic value of experience itself. Such people are likely to be more sensitive than others to both positive and negative experiences, and in previous studies openness has been shown to be related to both PAS and NAS, but not to affect balance (McCrae, 1983). Examination of Table 1 supports this hypothesis in men, and partially supports it among women. Openness to aesthetic experience is particularly associated with PAS; openness to fantasy with NAS. Openness to feelings, appropriately, is positively related to both, as is overall openness.

Avoiding Artifacts A Second Opinion on Personality

The relations of extraversion and neuroticism to subjective well-being are, from one point of view, not surprising. People who are chronically depressed and anxious are likely to be unhappy; those who are friendly, self-confident, and cheerful are likely to be happy. But on a topic where so many of our expectations are wrong, these basic relationships stand out. Andrews and Withey, for example, report that family life-cycle stage, age, income, education, race, and sex together account for only 8% of the variance in well-being. From Table 1 we can see that extraversion alone accounts for about the same amount of variance, and neuroticism accounts for as much as 27%. Personality is probably the strongest known predictor of well-being.

But Lawton (personal communication, April, 1983) has pointed out that there is a potential confound in these comparisons. In all our studies to date, self-reports of personality have been correlated with self-reports of well-being. Any biases that affect responses to one will probably also affect the other. This is particularly problematic in the case of Bradburn's scales, where key words--"anxious", "excited", "lonely", "proud"--often are also found in the items of personality inventories. The peculiarities of each individual's vocabulary may act across instruments to inflate the correlations. At a more fundamental level, both instruments draw on the individual's particular view of him- or herself and of the surrounding world, and it is always a matter of question how well individuals know themselves, or at least how accurate their self-reports are.

If we are interested in subjective well-being, of course, we have no choice but to ask the individual directly. If I claim to be unhappy, who can dispute it? Self-reports are virtually unavoidable in this context (Carp, 1977). But personality is different. Although inner thoughts and experiences are among the most important indicators of personality, they are by no means the only manifestations. Neuroticism, extraversion, and openness to experience ought to show up in overt behaviors, in interpersonal relationships, in the expression of emotions, in attitudes and values. An observer with sufficient familiarity should be able to rate an individual's personality even without access to inner experience, and the literature on the correspondence between self-reports and ratings confirms this expectation (McCrae, 1982).

Observer ratings are not infallible, either. Raters may be biased themselves, or poor observers, or prone to the kinds of response styles (like acquiesence) that may distort self-reports. But they do have the clear advantage of being an independent source of data. While response biases may inflate correlations between one self-report and another, they restrict correlations between self-reports and observer ratings by introducing unshared error variance. The correlations in Table 1 may represent the upper bound of the personality/well-being relationship; correlations with rated personality would provide a lower bound.

Table 2 provides just such correlations using ratings made by husbands and wives of subjects on a third-person version of the NEO Inventory (see McCrae, 1982, for details). Correlations are somewhat smaller than in self-report studies, but the pattern of results is strikingly similar. All the neuroticism facets are related to well-being, as are warmth, assertiveness, and positive emotions from the extraversion domain. Spouse-rated neuroticism appears to be a much better predictor than extraversion; even PAS is more strongly related to neuroticism than to extraversion. Perhaps the absence of joy is interpreted by external observers as a sign of maladjustment. The openness scales also show the same pattern seen in self-reports, though the correlation of total openness with NAS does not reach significance.

Judging from these correlations, facets of extraversion account for about 5% of the variance in well-being, and neuroticism for about 14%. Once again, these compare favorably with demographic predictors.

The Enduring Influence of Dispositions

The importance to adult development of the relation between personality and well-being stems from the fact that personality dispositions are extraordinarily stable in adulthood. A series of longitudinal studies (Block, 1977; Costa & McCrae, 1977; Costa, McCrae, & Arenberg, 1980; Leon, Gillum, Gillum & Gouze, 1979) have shown that objectively measured personality dispositions routinely show retest correlations of .70 or higher over intervals of six to thirty years, and that stability is found equally in young and old adults. In addition, studies on mean level differences in personality (Costa & McCrae, 1978; Douglas & Arenberg, 1978; Siegler, George, & Okun, 1979) have consistently shown little or no change in the average level of traits. As Dibner remarked, "At any point in time, a person is more like he has always been than he is like peers of his age group. A mature, nonneurotic younger person is likely to be a well-adjusted older person. The neurotic aged were most likely neurotic through much of their lives" (1975, p. 80).

To the extent that well-being depends on personality, it follows that an individual's well-being can be predicted years in advance by assessment of personality. Psychologists are not prophets, and we cannot predict whether life will hold wealth or poverty, health or illness, love or loss. But--if our model is correct--we can predict how individuals will evaluate

Table 2

Correlations Between Spouse Rated NEO Scales
And 1979 Well-Being Measures in Men and Women

NEO Scales	PAS	NAS	Balance	Satis-faction	D-T Scale
Anxiety	-23***	30***	-35***	-26***	-31***
Hostility	-17**	24***	-27***	-28***	-26***
Depression	-25***	38***	-41***	-29***	-35***
Self-Consciousness	-23***	24***	-31***	-16**	-28***
Impulsiveness	-10	22***	-21***	-29***	-19**
Vulnerability	-19**	29***	-32***	-21***	-36***
Neuroticism	-24***	36***	-40***	-32***	-37***
Warmth	21***	-06	18**	22***	20***
Gregariousness	12*	02	07	05	10
Assertiveness	10	-05	10	03	07
Activity	06	06	00	-03	04
Excitement Seeking	02	07	-03	-12*	-02
Positive Emotions	21***	-07	18**	15**	26***
Extraversion	19***	-01	13*	08	17**
Fantasy	04	20***	-10	-15*	-07
Aesthetics	15*	06	06	00	01
Feelings	05	19**	-09	-03	-02
Actions	09	01	05	04	06
Ideas	07	-07	10	02	11
Values	12*	05	04	03	06
Openness	14*	10	02	-02	04

Note: N = 296 for Bradburn scales; 291 for Satisfaction; 287 for D-T Scale.

Decimal points omitted. *p < .05; **p < .01; ***p < .001.

whatever life circumstances they encounter, whether they will be happy or unhappy with their lot (cf. Conley, in press).

This is a bold claim, but one supported by data. In the Boston longitudinal study, extraversion and neuroticism scores from the Sixteen Personality Factor Questionnaire (Cattell, Eber, & Tatsuoka, 1970) were significantly related to well-being measures administered ten years later (Costa & McCrae, 1980a). Scores from the Guilford-Zimmerman Temperament Survey (GZTS; Guilford, Zimmerman, and Guilford, 1976) predicted happiness, life satisfaction, and personal adjustment to aging over 6 and 12 year intervals in the BLSA sample (Costa, McCrae, & Norris, 1980). The well-being batteries administered in 1979 and 1981 provide another opportunity to test this prediction.

Table 3 gives the correlations between GZTS scores collected in the period from 1959 to 1969 with well-being data collected ten to 23 years later. The mean predictive interval for the first administration is 15.6 years; it is 17.7 years for the second. To interpret the findings, it is necessary to note that factor analyses (Costa & McCrae, in press) have shown that General Activity, Restraint (reversed), Ascendance, and Sociability load on a factor of extraversion; Emotional Stability, Objectivity, Friendliness (or low hostility), and Personal Relations (and to a lessor extent, Masculinity) define the opposite pole of a neuroticism factor. Thus, Table 3 shows that three of the extraversion scales--General Activity, Ascendance, and Sociability--are related to well-being, and especially PAS; all of the neuroticism scales except Masculinity are related to well-being, and especially NAS. These correlations are not markedly different in magnitude from contemporaneous measures, and provide a strong retrospective-predictive replication of the findings.

The GZTS Restaint scale (reversed) most closely resembles the Excitement Seeking facet of the NEO Inventory, and, like Excitement Seeking, is an aspect of extraversion unrelated to well-being. The GZTS Thoughtfulness scale loads on an openness to experience factor, and is related to PAS (in 1981) and NAS (1979), but never to global well-being. Despite a lapse of nearly two decades and the use of an entirely different instrument, even the details of the personality/ well-being model appear to be replicated.

Personality, it appears, is a lifelong determinant of subjective well-being. Individuals high in neuroticism are likely to see the problems of middle age as a "crisis" (Costa & McCrae, 1978); they will worry about increasingly poor health (Costa & McCrae, 1980b); they will be frustrated and disappointed by retirement, and are at risk for depression and Erikson's (1951) "despair" in old age. If, however, they are closed to experience, these affects may be blunted a bit; if they are extraverted, then their sorrows may be offset by the joy, warmth, and excitement that tend to accompany this disposition.

Table 3

Correlations Between GZTS Scales Administered 1959-1969
And Well-Being Measures from Two Administrations

GZTS Scales:	PAS	NAS	Balance	Satis-faction	D-T Scale
1979 Well-Being Administration					
General Activity	20**	-03	16*	11	04
Restraint	-06	-06	00	14*	02
Ascendance	18**	-11	19**	09	17*
Sociability	23***	-14*	25***	15*	23*
Emotional Stability	24***	-34***	39***	33***	34***
Objectivity	07	-23***	20**	28***	22***
Friendliness	10	-24***	22***	26***	18**
Thoughtfulness	05	15*	-07	-05	-05
Personal Relations	05	-27***	21**	17*	10
Masculinity	-06	-13*	04	08	05
1981 Well-Being Administration					
General Activity	18*	-13*	23**	18*	17*
Restraint	-05	-03	-02	08	07
Ascendance	28***	-13*	28***	17*	17*
Sociability	30***	-09	27***	20**	11
Emotional Stability	09	-17*	21*	33***	31***
Objectivity	03	-19*	16*	28***	22**
Friendliness	-04	-12	08	23**	20*
Thoughtfulness	23**	04	09	-01	05
Personal Relations	-07	-03	-02	14	05
Masculinity	-07	00	-04	04	04

Note: N = 214-239 for 1979 well-being administration; 158-182 for 1981 administration.

Decimal points omitted. *p < .05; **p < .01; ***p < .001.

Fortunately, most people are not high in neuroticism, and survey research shows that a full 35% consider themselves "very happy" (Gurin, Veroff, & Feld, 1960). Adjusted introverts are likely to meet the challenges of aging with equanimity; and adjusted extroverts will show the *joie de vivre* that make many older men and women an inspiration to the rest of us.

These are sweeping generalizations, and have a fatalistic quality that will probably be unwelcome. They do not bring words of encouragement about aging to those who need it most--the unhappy young and middle-aged. But they do appear to accurately reflect the facts, and those who would understand affect in adult development cannot afford to ignore them. We need to consider both the implications and the limitations of these generalizations in order to fully appreciate them.

Reassessing the Quality of Life in Old Age

Enduring personality dispositions have important effects on well-being, particularly since they operate year-in, year-out over the entire adult lifespan. On the other hand, at any given time personality variables appear to account for no more than a quarter of the variance in well-being. Even allowing for unreliability of measurement, it is clear that a substantial portion of variance must be accounted for by something else. We have pointed out the classic error of assuming that subjective well-being mirrors objective life quality; we must also warn against the danger of dismissing well-being as nothing but chronic complaining or groundless optimism.

What does it *mean* when an old woman tells us she is perfectly happy living alone, or an old man complains that local transportation systems do not meet his needs? Do these comments reflect real assets or deficiencies in their environments? Should they be taken at face value, and used as the basis for interventions to enhance life quality? Do they represent only recent and relative changes in quality of life, as adaptation theory would suggest? If so, different evaluations could be expected as time passes. Or are they chiefly reflections of longstanding dispositions, equally impervious to the assaults of an unsympathetic world, and to the assistance of benevolent social institutions? Until research clarifies these issues, we will be on shaky grounds in employing only subjective well-being as the criterion.

One alternative is to disregard well-being entirely. Rosow (1977) has argued that subjective well-being is essentially irrelevant as a basis for social policy decisions: we ought to improve the quality of life for individuals whether they appreciate it or not. There is much to be said for this position, but its implementation depends on our ability to identify aspects of the good life. More importantly, given limited resources, we must be able to prioritize them, but we have no guarantees that our priorities correspond to those of older individuals. And how, without consulting him or her, can we design programs best suited to the specific needs of each individual?

There must be a middle ground between uncritical acceptance and complete disregard of subjective judgments of satisfaction. Several possibilities suggest themselves. Responses to individual items in a life satisfaction index may represent the joint influence of external circumstances and personal dispositions; when added together, the external circumstances may vary and cancel out, while the personal dispositions will emerge as the main determinant of total score. The influence of objective circumstances is likely to be most pronounced when specific areas of satisfaction or dissatisfaction are addressed. For example, satisfaction with health declines with age, even though overall satisfaction does not (Campbell, Converse, & Rodgers, 1976). An intervention to improve the diet of older people ought to result in improved satisfaction with food, and possibly with health, but probably would not noticeably affect global morale (cf. Carp, 1977).

Even specific judgments, however, are made against the individual's subjective standards, and these are influenced to an unknown degree by processes of adaptation and by characteristic levels of optimism or pessimism. It might be more informative to ask individuals to evaluate their life quality against other objective situations. The wealthy are not much happier than the poor, but, given the choice, we'd rather be rich, and that choice is perhaps a better indicator of the value of money than is any rating of satisfaction. Individuals may be better able to evaluate the quality of life and suggest ways of improving it through such comparisons.

Another Look at Old Age

Discussions of adaptation level and of the influence of personality on satisfaction easily lead to the impression that subjective judgments are narcotics that lull us into accepting the inevitable. This chapter may have reinforced the prevalent notion that life after sixty really is miserable, even though older people may feel satisfied. Such an impression would be unwarranted.

No one would deny that the elderly in America face a number of stresses. Often economic resources are limited, with little prospect of improvement. Friends and relatives die, health and vigor decline, memory begins to fail. But no age is without its drawbacks, and in many respects, the ills of old age have been exaggerated.

For example, most older people have frequent and rewarding contact with younger family members (Troll, Miller, & Atchley, 1979). They enjoy the blessings of grandparenthood without the responsibilities of parenthood. As most gerontologists--but few laypersons--know, only about 5% of those over 65 are institutionalized, with most older Americans living in their own homes. Retirement, once thought to be enforced idleness, is enjoyed by many individuals as a time for leisure and as much or little productive activity as they care for. If and when we devise an absolute measure of life quality independent of the subjective judgment of the individual and its personality determinants, we may well find that old age is the happiest time of life.

References

Andrews, F. M., & Withey, S. B. Social indicators of well-being: Americans' perception of life quality. New York: Plenum, 1976.

Block, J. Advancing the psychology of personality: Paradigmatic shift or improving the quality of research. In D. Magnusson & N.S. Endler (Eds.), Personality at the cross-roads: Current issues in interactional psychology. Hillsdale, N.J.: Lawrence Erlbaum, Associates, 1977.

Bradburn, N. M. The structure of psychological well-being. Chicago: Aldine, 1969.

Brickman, P., & Campbell, D. T. Hedonic relativism and planning the good society. In M. H. Appley (Ed.), Adaptation level theory: A symposium. New York: Academic Press, 1971.

Brickman, P., Coates, D., & Janoff-Bulman, R. Lottery winners and accident victims: Is happiness relative? Journal of Personality and Social Psychology, 1978, 36, 917-927.

Campbell, A, Converse, P. E., & Rodgers, W. L. The quality of American life: Perceptions, evaluations, and satisfactions. New York: Russell Sage Foundation, 1976.

Carp, F. M. Morale: What questions are we asking of whom? In C. N. Nydegger (Ed.), Measuring morale: A guide to effective assessment. Washington, DC: Gerontological Society, 1977.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. The handbook for the sixteen personality factor questionnaire. Champaign, IL: Institute for Personality and Ability Testing, 1970.

Conley, J. J. The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality, and self-opinion. Personality and Individual Differences, in press.

Costa, P. T., Jr., & McCrae, R. R. Age differences in personality structure revisited: Studies in validity, stability, and change. Aging and Human Development, 1977, 8, 261-275.

Costa, P. T., Jr., & McCrae, R. R. Objective personality assessment. In M. Storandt, I. C. Siegler, & M. F. Elias (Eds.), The clinical psychology of aging. New York: Plenum Press, 1978.

Costa, P. T., Jr., & McCrae, R. R. Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. Journal of Personality and Social Psychology, 1980, 38, 668-678. (a)

Costa, P. T., Jr., & McCrae, R. R. Somatic complaints in males as a function of age and neuroticism: A longitudinal analysis. Journal of Behavioral Medicine, 1980, 3, 245-257. (b)

Costa, P. T., Jr., & McCrae, R. R. Still stable after all these years: Personality as a key to some issues in adulthood and old age. In P. B. Baltes & O. G. Brim (Eds.), Life span development and behavior, (Vol. 3). New York: Academic Press, 1980. (c)

Costa, P. T., Jr., & McCrae, R. R. Stress, smoking, and psychological well-being: The illusory benefits of smoking. Advances in Behavior Research and Therapy, 1981, 3, 125-150.

Costa, P. T., Jr., & McCrae, R. R. Concurrent validation after 20 years: Implications of personality stability for its assessment. In J. N. Butcher & C. D. Spielberger (Eds.), Advances in Personality Assessment, Vol. 4. Hillsdale, NJ: Erlbaum, in press.

Costa, P. T., Jr., McCrae, R. R., & Arenberg, D. Enduring dispositions in adult males. Journal of Personality and Social Psychology, 1980, 38, 793-800.

Costa, P. T., McCrae, R. R., & Norris, A. H. Personal adjustment to aging: Longitudinal prediction from Neuroticism and Extraversion. Journal of Gerontology, 1981, 36, 78-85.

Dibner, A. S. The psychology of normal aging. In M. G. Spencer & C. J. Dorr (Eds.), Understanding aging: A multidisciplinary approach. New York: Appleton-Century-Crofts, 1975.

Dicken, C. Good impression, social desirability, and acquiescence as suppressor variables. Educational and Psychological Measurement, 1963, 23, 699-720.

Douglas, K., & Arenberg, D. Age changes, cohort differences, and cultural change on the Guilford-Zimmerman temperament survey. Journal of Gerontology, 1978, 33, 737-747.

Erikson, E. H. Childhood and society. New York: Norton, 1950.

Eysenck, S. B. G., & Eysenck, H. J. On the dual nature of extraversion. In Personality Structure and Measurement, London: Routledge and Kegan Paul, 1969.

Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. The Guilford-Zimmerman Temperament Survey Handbook: Twenty-five years of research and application. San Diego, CA.: EdITS Publishers, 1976.

Gurin, G., Veroff, J., & Feld, S. Americans view their mental health. New York: Basic Books, 1960.

Helson, H. *Adaptation-level theory*. New York: Harper & Row, 1964.

Herzberg, F., Mausner, B., & Snyderman, B. B. *The motivation to work*. New York: Wiley, 1959.

Herzog, A. R., Rodgers, W. L., & Woodworth, J. *Subjective well-being among different age groups*. Ann Arbor: Insitute for Social Research, University of Michigan, 1982.

Lawton, M. P. The varieties of well-being. In C. Malatesta & C. Izard (Eds.), *Affective processes in adult development and aging*. Beverley Hills, CA: Sage, in press.

Leon, G. R., Gillum, B., Gillum, R., & Gouze, M. Personality stability and change over a 30 year period-middle age to old age. Journal of Consulting and Clinial Psychology, 1979, *23*, 245-259.

McCrae, R. R. Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology*, 1982, *43*, 293-303.

McCrae, R. R. Extraversion is not a filter, neuroticism is not an outcome: A reply to Lawton. *Experimental Aging Research*, 1983, *9*, 73-76.

McCrae, R. R., & Costa, P. T., Jr. Joint factors in self-reports and ratings: Neuroticism, extraversion, and openness to experience. *Personality and Individual Differences*, 1983, *4*, 245-255.

McCrae, R. R., & Costa, P. T., Jr. Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 1983, *51*, 882-888.

Palmore, E., & Kivett, V. Change in life satisfaction: A longitudinal study of persons aged 46-70. *Journal of Gerontology*, 1977, *32*, 311-316.

Rosow, I. Morale: Concept and measurement. In C. N. Nydegger (Ed.), *Measuring morale: A guide to effective assessment*, Washington, DC: The Gerontological Society, 1977.

Siegler, I. C., George, L. K., & Okun, M. A. Cross-sequential analysis of adult personality. *Developmental Psychology*, 1979, *15*, 350-351.

Tamir, L. M. *Men in their forties: The transition to middle age*, New York: Springer, 1982.

Troll, L. E., Miller, S. J., & Atchley, R. C. *Families in later life*. Belmont, CA: Wadsworth, 1979.

Warr, P. Barter, J., & Brownbridge, G. On the independence of positive and negative affect. Journal of Personality and Social Psychology, 1983, 44, 644-651.

Zuckerman, M. Sensation seeking: Beyond the optimal level of arousal. New York: Lawrence Erlbaum Associates, 1979.

Elias, M.F., and Streeten, D.H.P. (Eds.),
Hypertension and Cognitive Processes.
Mt. Desert, Me.: Beech Hill, 1980

Hypertension, Somatic Complaints, and Personality

Paul T. Costa, Jr., Ph.D., Robert R. McCrae, Ph.D.,
Reuben Andres, M.D., and Jordan D. Tobin, M.D.

Hypertension is often classed as a psychosomatic illness [1; 47] on the premise that psychological factors form one of the major determinants of the disease. In part, this premise is based on the established findings that chronic stress can lead to the development of hypertension [35].

DOES PERSONALITY CAUSE HYPERTENSION?

Clinical experience has led many to the view that hypertension or elevated blood pressure is related to personality; specifically, that it may result from the chronic internal stress or conflict over repressed rage or anger [10; 25]. These theoretical positions are supported by at least one epidemiological study [16] which finds an association between suppressed anger and hypertension. In a number of studies using standard measures such as the MMPI [23], Maudsley Personality Inventory [37], and the Sixteen Personality Factors Questionnaire [20], investigators have reported that individuals under treatment for hypertension score higher on measures of maladjustment or neuroticism. Studies using the Cornell Medical Index (CMI) have indicated higher reporting of symptoms among hypertensives [27; 30; 49], perhaps indicating hypochondriasis or excessive body concern. Research using interview techniques or clinical ratings have also generally found evidence of psychological disturbance [8; 13].

These studies do not necessarily support the hypothesis that anger or "repressed rage" is a unique feature in the personality profiles of hypertensives, but they can be seen as evidence that individuals treated for hypertension are more likely to suffer from general maladjustment, emotional instability or neuroticism, of which hostility is one aspect.

This interpretation of studies on personality factors in hypertension has not gone unchallenged. Two powerful rival hypotheses have been offered which might account for the pattern of findings. Individuals higher in neuroticism (who are likely to manifest a variety of maladaptive traits, including anxiety, hostility, and depression as well as hypochondriasis) are more likely to visit a doctor [45], be diagnosed as hypertensive, and be participants in clinical studies of hypertension [17; 18]. In fact, when studies have examined the relation between neuroticism and blood pressure [32] in unselected, representative national samples, no association with hypertension was found. If some aspect of neuroticism caused hypertension, we would expect to find an association between these variables in non-clinic-selected populations; the failure to find significant relations here suggests that neuroticism leads only to the *discovery* or *detection* of hypertension. Not only do more neurotic individuals see physicians more frequently, they also give complaints of headaches and dizziness which lead physicians to check blood pressure [8]. The disproportionate discovery of hypertension among neurotic individuals may lead to spurious associations between these variables in clinical studies.

The second rival explanation holds that individuals who are told that they have hypertension and are warned of the seriousness of the illness, may become more anxious, depressed, or hostile as a result. They may manifest increases in hypochondriacal bodily concern as they look for symptomatic evidence of their condition. This hypothesis has been entertained with increasing frequency in the literature [8; 20; 24; 48]. For example, Stewart and Lond [42] in a study of 200 hypertensive patients found that most hypertensives (71 out of 96) who knew their diagnosis complained of non-specific headaches, while most (87 out of 104) of those who were unaware of their diagnosis reported no headache. The same trend was present even in patients with malignant hypertension. Of 12 knowing their diagnosis, 10 had headaches compared with 1 of 6 patients who were ignorant of their hypertension. Although certain kinds of headache (i.e., severe, of recent origin, present on waking and often associated with nausea and progressive impairment of vision) may be a physiological effect of severe or malignant hypertension [21], many of the nonspecific types of headache complaint appear instead to be unrelated to physiological status.

DOES HYPERTENSION AFFECT PERSONALITY?

For those interested in the psychosomatic hypothesis, these considerations represent artifacts. But it is possible to view them as tentative evidence for another hypothesis, namely, that hypertension may be not the effect, but the cause of changes in personality. Herbert Weiner [47], in reviewing the current status of psychosomatic research on essential hypertension, concludes that patients with hypertension are psychologically heterogeneous, just as essential hypertension itself is not a single disease entity. He writes, "Even if the psychological characteristics of these patients were uniform, it might not have etiological or pathogenetic significance. Because these characteristics are observed after the onset of the disease, they might be a product of high blood pressure levels or high serum renin and angiotensin activity (p. 113)."

Weiner's hypothesis encounters many of the same problems as the older one, but it has the advantage of explaining why prospective longitudinal studies [39; 46] fail to find predictive associations between aspects of personality and hypertension. If longitudinal data included baseline measurements of blood pressure and subsequent measurements of personality, the hypothesis could be directly tested.

Wood, et al. [48] reported that essential hypertensive patients had higher scores than normotensive controls on both physical and psychiatric sections of the Cornell Medical Index (CMI). However, the authors caution that these results indicate only that "hypertensive subjects that *knew* they were hypertensive *reported* more physical and psychological symptoms" (p. 428). In a later study [49], the same authors examined the effects of diagnosed uncomplicated essential hypertension on the Zung Depression and the Spielberger State and Trait Anxiety scales. Spielberger [41] has distinguished between state anxiety, a transitory emotional state, and trait anxiety, a stable individual difference in anxiety-proneness. Hypertensives, particularly the younger men, showed significantly higher scores than controls on the Depression Scale and the State Anxiety scale. They were *not* significantly higher on Trait Anxiety. Anxiety for these subjects appears to be a temporary, situationally-induced reaction or state, not a long-standing disposition or trait.

The data from the Wood, et al. [49] study do not directly address the question of whether the cause of the anxiety is hypertension itself or the diagnosis and treatment of it; nor whether the anxiety state signals the beginning of a permanent change in personality which might be characterized as a trait some years later.

BLOOD PRESSURE, PERSONALITY, AND MEDICAL COMPLAINTS IN THE BLSA MEN

In an attempt to answer these kinds of questions, data on systolic and diastolic blood pressure, the CMI, and the Guilford-Zimmerman Temperament Survey [GZTS; 15] were examined among more than 700 male subjects in the Baltimore Longitudinal Study of Aging.

Participants in the Baltimore Longitudinal Study are a community-dwelling, generally healthy group of male voluteers, 96% white, who have agreed to return for testing at fixed intervals. The majority (80%) work in or are retired from scientific, professional, or managerial positions. Most (93%) are high school graduates, and 71% are college graduates; 88% were married. Data in this paper are from responses of participants who entered the study from late 1958 through 1978. At the time of first administration of the GZTS, age ranged from 17 to 97 ($N = 796$, mean = 49.8 years).

Second administration data were obtained from over 500 men age 25 to 91. The smaller sample size is due to the varying number of years in the study as well as death and withdrawal. Subjects who returned for the second or third administration tended to be higher than non-repeats in Emotional Stability, Objectivity, Friendliness, and Personal Relations, and lower in Ascendance [9].

Most importantly, from the viewpoint of this chapter, the subjects were not a clinical population. Although volunteers, and not fully representative of the male population of the United States, they are relatively free from the problem of self-selection for medical complaints. Evidence of this is seen in the generally low CMI scores. Since the BLSA does not provide treatment, medical information may be perceived by the subject in a somewhat different context.

A number of subjects in this study ($N = 101$) were on medication for hypertension either upon entry into the study or at the succeeding four visits. Because being on medication may increase somatic concern while decreasing blood pressure, including these subjects would obscure any real positive association between those two variables. Accordingly, these subjects were omitted from all analyses. No other screening was done in the major analyses to be reported.

Single blood pressure recordings are known to be unstable or variable [11]. In order to provide a more stable estimate of these variables, four systolic and four diastolic readings taken over a two day period by the nursing staff were averaged. Table 1 shows the intercorrelation of these four readings. These values are similar to the average week-to-week correlations of .62 for systolic pressure recorded by health professionals in a sample of mild hypertensives studied by Engel and his associates [11]. The internal consistency (coefficient alpha) of the average systolic and diastolic pressures were .92 and .86 ($N = 761$) respectively. These readings were taken as part of a basal metabolism rate procedure. Subjects were resting and fasting, and readings were taken early in the morning. In consequence, the values reported are about 10 mm Hg lower than the casual sitting pressures recorded by physicians during physical examinations. Average systolic

Table 1

Intercorrelation of Four Supine Blood Pressure Readings Over a Two-Day Period

	First Day	Second Day	
	Reading 2	Reading 3	Reading 4
Systolic ($N = 763$-829)			
First Day			
Reading 1	.93	.72	.73
Reading 2		.70	.72
Second Day			
Reading 3			.92
Diastolic ($N = 760$-828)			
First Day			
Reading 1	.86	.51	.50
Reading 2		.54	.53
Second Day			
Reading 3			.83

pressure ranged from 89 to 189 (mean = 119.0, *SD* = 14.8). Average diastolic pressure ranged from 45 to 113 (mean = 75.2, *SD* = 8.9). About 9% of the sample (74 men) had systolic pressure above 140; about 6% (51 men) had diastolic pressure above 90.

Each subject was given the standard GZTS and CMI instructions individually and completed the questionnaire during the remainder of his first or second 3-day visit to the Gerontology Research Center. Subjects were re-administered the GZTS approximately every six years. Because of complications in scheduling, a few subjects took the test two years in succession. To maintain consistency of the time interval, longitudinal analyses are limited to subjects who took their second GZTS 4 to 12 years after their first medical examination. The CMI was re-administered on the fifth visit;

Table 2

Contemporaneous Relations Between
Blood Pressure, CMI Sums, and GZTS Scales

	Systolic	Diastolic
CMI ($N = 723\text{-}731$)		
Physical (A-L) Sum	.18***	.12**
Psychiatric (M-R) Sum	-.02	-.01
GZTS ($N = 661\text{-}724$)		
General Activity	-.09*	-.03
Restraint	.14***	.10**
Ascendance	-.11**	-.07
Sociability	-.04	-.05
Emotional Stability	.04	.01
Objectivity	.06	.06
Friendliness	.11**	.09*
Thoughtfulness	.05	.02
Personal Relations	.06	.08*
Masculinity	-.12**	-.09*

*$p < .05$
**$p < .01$
***$p < .001$

longitudinal analyses again are limited to re-tests taken 4 to 12 years after the first blood pressure readings. The average predictive intervals were 7.4 years for the GZTS and 6.5 years for the CMI.

Table 2 gives the correlations of personality (GZTS) and physical and psychiatric scores of the CMI with average systolic and diastolic pressure. Eleven of these 24 correlations are statistically significant, though all are quite small. It would appear that higher levels of blood pressure are associated with higher somatic complaints, Restraint, Friendliness, and good Personal Relations, and with lower General Activity and Masculinity.

Although age and GZTS personality scales show only small correlations, age is clearly related to total CMI physical complaints, and substantially correlated with blood pressure. In fact, the correlations in the present sample are .56 for systolic and .38 for diastolic pressure. Table 3 presents the

Table 3

Contemporaneous Relations Between Blood Pressure, CMI Sums, and GZTS Scales Controlling for Age

	Systolic	Diastolic
CMI (*N*=720-728)		
Physical (A-L) Sum	.05	.03
Psychiatric (M-R) Sum	.01	.01
GZTS (*N*=658-721)		
General Activity	.03	.05
Restraint	-.04	-.02
Ascendance	.02	.01
Sociability	.02	-.02
Emotional Stability	.02	.00
Objectivity	.05	.05
Friendliness	-.01	.01
Thoughtfulness	.01	-.01
Personal Relations	-.06	.01
Masculinity	-.02	-.03

same correlations as those shown in Table 2 but with the effects of age partialled out. None of these correlations is statistically significant. Given the reliability of both personality and blood pressure measures and the size and nature of the sample, this is rather clear evidence of the *independence* or lack of association between personality and blood pressure.

The traits measured by the GZTS have been shown to be stable over time [6], and reliable, averaged measurements of blood pressure are also relatively stable. Between the first and fifth medical examinations (from 4 to 8 years) the stability coefficient of average systolic pressure was .66 ($N=288$, $p<.001$); of average diastolic, .50 ($N=290$, $p<.001$). Although contemporaneous, both sets of variables in Table 2 describe relatively enduring conditions or characteristics of the individual. If elevated blood pressure itself affected personality, it should be evident in these correlations. However, it is possible that the temporal interval between an increase of blood pressure and the measurement of personality was not sufficient.

Table 4

Predictive Relations Between Blood Pressure and CMI Sums and GZTS Scales Controlling for Age

	Simple Correlations		Partial Correlations	
	Systolic	Diastolic	Systolic	Diastolic
CMI ($N = 410$-417)				
Physical(A-L) Sum	.10*	.12*	.02	.08
Psychiatric(M-R) Sum	-.01	.05	.00	.06
GZTS ($N = 345$-376)				
General Activity	-.02	-.02	.09	.05
Restraint	.05	.02	-.08	-.08
Ascendance	.00	.02	.11*	.10
Sociability	.08	.07	.12*	.09
Emotional Stability	.10	.01	.07	-.02
Objectivity	.05	.01	.01	-.01
Friendliness	.00	-.03	-.08	-.09
Thoughtfulness	.01	.04	-.03	.01
Personal Relations	-.10	-.10	-.15**	-.12*
Masculinity	-.12*	-.14**	-.06	-.09

Note: Predictive interval = 4-12 years, mean intervals: 6.5 years for CMI, 7.4 years for GZTS.
*$p < .05$
**$p < .01$

In order to test this possibility, average pressures at first measurement were correlated with GZTS and CMI scores collected 4 to 12 years later. Table 4 presents these cross-lagged correlations, again with age statistically controlled. Here, systolic pressure is correlated with increased Ascendance and Sociability and poorer Personal Relations. If chronic hypertension has any effect on these dimensions of personality, it is so small as to be of no practical consequence. However, these correlations are only marginally significant, and are probably best interpreted as chance findings.

Certainly some might object that the traits measured by the GZTS, a self-report instrument, are not equivalent to the dispositions or personality characteristics thought to be related to hypertension. For example, individuals, almost by definition, are not aware of *repressed* rage and could not accurately attribute that trait to themselves. A clinical judgment is usually required, though such judgments tend to be quite unreliable. Given the available data, we attempted to approximate these clinical judgments by examining patterned combinations of traits. It is, after all, on the basis of such a pattern of responses from the subject that clinicians make their judgments.

To do this, we performed analyses of variance in which we cross-classified individuals on the basis of trichotomized personality scale scores and looked for significant interactions between the traits. Specifically, using scales F (Friendly vs. Hostile), R (Restrained vs. Impulsive), A (Ascendant vs. Submissive), and M (Masculine vs. Feminine), we tested 6 hypotheses that higher average pressure might be found in individuals who were either (1) *hostile but restrained*, (2) *ascendant but restrained*, (3) *masculine but restrained*, (4) *hostile but submissive*, (5) *masculine but submissive*, (6) *hostile but feminine*. Neither main effects nor interactions proved significant when age was used as a covariate.

Other analyses were also conducted. Scatterplots of average pressure by each GZTS and CMI variable were examined to detect a possible curvilinear relationship. None was observed. In order to eliminate possible artifacts introduced by illness or other drugs, the analyses presented in Tables 2-4 were re-done using a physician's reading of blood pressure on a subsample of drug- and disease-free individuals. Three of the twenty-four contemporaneous correlates were significant with age covaried, but none of these accounted for even as much as 1% of the variance.

Finally, an attempt was made to address the conception of hypertension as a disease process, distinct from normal variation in blood pressure. This requires the selection of a cut-point above which individuals are considered hypertensive. The choice of such a point is both arbitrary and controversial. We decided to take diverse groups to minimize the number of false positives. For these analyses, hypertension was defined as average systolic pressure of 150 to 189 or average diastolic pressure of 93 to 113. Hypotension was defined as systolic from 89 to 100 or diastolic from 45 to 60. A normotensive group was also included with systolic pressure at 119-120 and diastolic pressure at 74-75. Analyses of covariance with age as the covariate showed no effects for hypertension classification on any of the personality scales or the CMI sums.

HYPERTENSION, PERSONALITY, AND ADAPTATION

Hypertension is known as a "silent" disease—during the many years it persists, it usually does not show itself to the individual through any percievable symptoms. It seems apparent that it also does not manifest itself in personality change which might be apparent to those around the individual. Conscientious individuals who maintain their health through regular check-ups and more or less neurotic individuals who bring repeated complaints to their physicians are likely to be diagnosed as hypertensive; but we find little evidence in the literature or our own data that the disease itself is related to neuroticism or any other objectively measurable personality trait.

Individuals who are informed that they are hypertensive often react by temporary anxiety, depression, or somatic concern—understandable responses in the circumstances. It appears, however, that these reactions are indeed temporary, with no permanent or long-term effects, as shown by the lack of predictive relations between blood pressure and personality and perceived health scores taken several years later. A dissertation study [28] which investigated the possible adverse psychological effects of treatment for hypertension similarly found little change over even so short an interval as six months.

Although consistent with most of the empirical literature [43], these findings are not supportive of some clinical conceptions nor of the psychosomatic hypothesis in general. They can be assimilated, however, to other established bodies of psychological knowledge. They reaffirm, for example, the conclusion that personality traits (and perceived health insofar as it is a function of personality) are stable in adulthood [6]. They may also be interpreted as instances of the process of adaptation through which individuals grow accustomed to situations which initially provoke emotional reactions.

The capacity to adapt to disturbing circumstances occasionally has paradoxical consequences. It reduces the psychological stress of unpleasant emotions, but does so at some cost. In the present instance, adaptational processes may be responsible for the phenomenon known as "threat inoculation". In fact, extensive research on the determinants of compliance with medical regimens has provided very mixed evidence on the value of fear as a motivator. On the other hand, the same body of research does support the usefulness of certain other practices. The treating physician should demonstrate concern for the patient, and should provide realistic information on the nature and danger of the disease, as well as on the effectiveness of treatment. Practicioners should realize that faithful adherence is not likely to happen spontaneously.

IMPLICATIONS FOR FUTURE RESEARCH AND TREATMENT

No issue is ever definitively settled in inductive sciences, and null results—the failure to find an association—are particularly susceptible to later revision. Some critics would argue that the objective personality inventories used in this and much other empirical research are insensitive to the deeper layers of personality which may be related to hypertension. Perhaps only projective tests or clinical ratings or interviews can elicit the necessary information. While this possibility cannot be gainsaid, it does not appear promising. The dangers of subjectivity, bias and unreliability in these methods seem to outweigh the limited likelihood that a real relationship exists which has gone undetected by objective measures. Certainly anyone wishing to pursue the hypothesis that personality influences the development of hypertension (or vice-versa) with these methods should use extreme care in designing and conducting research: patients and controls should be randomly selected or be carefully matched; judges should be blind to diagnosis; the reliability of ratings should be established; the fact and extent of the patients' knowledge of the condition (as well as his or her medications) should be carefully ascertained and used in analyzing data. Although the data presented here concerned blood pressure per se and personality, many would rightly insist that clinical hypertension be distinguished from temporary elevation, and that the types of hypertension ("borderline", essential, renovascular, malignant, etc.) and the stages of the disease be specified. In order to be credible, research would have to be done on large samples (several hundred subjects), with clear hypotheses formulated in advance of the research, and any significant results replicated on other samples. In order to distinguish transient from long-term effects, follow-ups of at least six months would be needed. Research which fails to meet all or most of these requirements will not contribute to our understanding of hypertension and its effects.

Similar conclusions and arguments have been advanced by others though from different considerations and prospectives. Adrian Ostfeld [29], in an invited editorial of *Psychosomatic Medicine*, asked whether the research since 1939 on unique personality types or specified intrapsychic conflicts in the hypertensive person "has led to an understanding of the cause and pathogenesis of high blood pressure or to better treatment of it." His answer is clearly no. He suggests abandoning such questions and focusing on others. Among the questions in the hypertension and personality field which it appears useful to pursue are: (1) What kinds of cognitive and behavioral treatments are effective in lowering blood pressure? (2) Which types of intervention are most effective for individuals characterized by different personality traits? (3) How can information on personality be used in the treatment of distress-related medical complaints among hypertensives? (4) How do personality-related differences in the utilization of medical care affect programs for the prevention of hypertension? (5) What personality factors affect compliance or non-compliance with prescribed medical treatment?

The independence of blood pressure from enduring personality dipositions certainly does not mean that physicians and psychologists have nothing further to offer each other in this area. It means, rather, that the focus should shift from *etiology* to *treatment*. Psychological factors such as stress can elevate blood pressure in those who are predisposed to that reaction—although it appears that physiological rather than personality traits mediate this effect [38; 40]. Conversely, therapies which tend to counteract this stress-response, like relaxation training [19; 14] and biofeedback [27] are promising approaches to treatment.

Of interest to the student of personality is the issue of personality-treatment interactions. Qualls and Sheehan [31] have reported that individuals open to absorbing experiences may benefit more from relaxation, while closed individuals profit from a biofeedback approach. Because these two techniques are promising approaches to the management of borderline hypertension, these results merit particular attention. Sixteen high absorption and 16 low absorption female subjects, defined by scores on an abbreviated version of Tellegen and Atkinson's [44] Absorption Scale, underwent a biofeedback and a no-feedback session with the order of conditions counterbalanced. Electromyographic (EMG) reductions were greater during no-feedback than during feedback for high absorption women; no differences across expermental conditions were found for low absorption women. In discussing these results, the authors focus on the interactions between subject's characteristic level of absorption and specific situational factors. Experimental conditions that facilitate or allow withdrawal from the external environment (relaxation) were most beneficial for persons with high capacity for absorbed attention; whereas for subjects low or limited in their capacity for absorbed attention, conditions that place an external, attentional demand on subjects, (biofeedback), were more effective. Conceptually, absorption can be seen as an aspect of the larger domain of openness to experience [6]. The role of traits in the three major personality domains of neuroticism, extraversion, and openness should all be investigated in relation to the treatment of hypertension.

Some evidence in regard to neuroticism has already been gathered. Bulpitt and others [5] report that *within* hypertensive populations, anxiety and depression are associated with a higher incidence of symptom complaints, and Davies [8] reviews studies tending to show that placebos alleviate these symptoms as effectively as standard pharmacotherapy for high blood pressure. It is worth noting that, regardless of the physiological status of these complaints, the individual does indeed suffer psychologically because of them. Treatment of the hypertension should be seen as the treatment of the hypertensive *person* and emotional support may be as important as medication to the afflicted individual.

The self-selection of subjects high in neuroticism is a problem for research on personality and hypertension; the relative neglect of medical attention among well-adjusted individuals is a problem for preventive medicine. Bergland and others [47], in a study which contrasted all the treated with all the undiagnosed 50-year old male hypertensives in Goteborg, Sweden, reported that the undiagnosed were higher in the need for autonomy and lower in symptom reporting. They conclude that hypertensive screening examinations may be necessary in order to identify those individuals who are unlikely to seek treatment themselves. In this case, it appears that the chronic distress of more neurotic individuals may work to their advantage, because it leads to a higher utilization of medical services.

Finally, one of the major problems in treating hypertension is compliance with medical regimens [22; 36]. Patients who are supposed to maintain a life-long program of anti-hypertension medication too often neglect to take medication, or give up the program when their blood pressure returns to normal levels. Research on compliance has often centered around aspects of the health beliefs model [3; 33; 34], which holds that certain attitudes toward health, illness, and medical care determine health behaviors. Future research should also include measurement of personality variables which may contribute to an understanding of the reasons for non-compliance, and aid in the design of individually-tailored interventions.

SUMMARY

Certain personality variables (e.g., repressed rage) have traditionally been thought to be among the determinants of hypertension. In recent years, however, critics have noted a number of problems with the methodology used in earlier studies, and better-controlled research has generally failed to find an association between personality and hypertension.

As an alternative, it has been proposed that hypertension may be responsible for changes in personality. Some evidence suggests that hypertension patients become more anxious, depressed, or concerned with bodily symptoms, although this may be an effect of knowledge of their condition or of treatment rather than of the hypertension itself, and it may be only a temporary change.

Data were offered on a large sample of adult males who had been followed over a period of years. When the influence of age was controlled, no association was found between blood pressure and any of twelve personality and psychosomatic concern measures; nor did patterned combinations of traits (e.g., hostile but restrained) show any relation. Attempts to predict personality scores four to twelve years later from blood pressure levels also failed, suggesting that any change in anxiety or somatic concern due to hypertension or its treatment are short-lived.

These conclusions prompted the recommedation that future research pursue other issues including: (1) the kinds of cognitive and behavioral treatments which may be effective; (2) the most appropriate interventions for individuals characterized by different personality traits (3) the use of personality information in the treatment of distress-related complaints among hypertensives; (4) the relation of personality variables to the utilization of preventive medical programs; and (5) the personality factors which affect compliance or non-compliance with medical regimens.

REFERENCES

1. Alexander, F. *Psychosomatic medicine.* New York: Norton, 1950

2. Basmajian, J. V. (Ed.) *Biofeedback: Principles and practice for clinicians.*Baltimore: Williams & Wilkins, 1979.

3. Becker, M. H. (Ed.) The health belief model and personal behavior. *Health Education Monograph,* 1974, *2,* (Whole No. 4).

4. Berglund, G., Ander, S., Lindstrom, B., & Tibblin, G. Personality and reporting of symptoms in normo- and hypertensive 50 year old males. *Journal of Psychosomatic Research,* 1975, *19,* 139-145.

5. Bulpitt, C. J., Dollery, C. T., & Hoffbrand, B. I. The contribution of psychological features to the symptoms of treated hypertensive patients. *Psychological Medicine,* 1977, *7,* 661-665.

6. Costa, P. T., Jr., & McCrae, R. R. Still stable after all these years: Personality as a key to some issues in adulthood and old age. In P. B. Baltes & O. G. Brim (Eds.), *Life span development and behavior,* (Volume III). New York: Academic Press, in press.

7. Costa, P. T., Jr., McCrae, R. R., & Arenberg, D. Enduring dispositions in adult males. *Journal of Personality and Social Psychology,* 1980, *38,* 793-800

8. Davies, M. H. Is high blood pressure a psychosomatic disorder? A critical review of the evidence. *Journal of Chronic Diseases,* 1971, *24,* 239-258.

9. Douglas, K., & Arenberg, D. Age changes, cohort differences, and cultural change on the Guilford-Zimmerman temperament survey. *Journal of Gerontology,* 1978, *33,* 737-747.

10. Dunbar, H. F. *Emotions and bodily changes.* New York: Columbia University Press, 1954.

11. Engel, B. T. Gaarder, K. R., & Glasgow, M. S. Behavioral treatment of high blood pressure: I. Intra- and interdaily variations of blood pressure during a one-month baseline period. In preparation.

12. Engel, B. T. Personal communication, January 25, 1980.

13. Friedman, M. J., & Bennet, P. L. Depression and hypertension. *Psychosomatic Medicine*, 1977, *39*, 114.

14. Goldfried, M. R. The use of relaxation and cognitive relabeling as coping skills. In R. B. Stuart (Ed.), *Behavioral self-management*. New York: Brunner/Mazel, 1977.

15. Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. *The Guilford-Zimmerman temperament survey handbook: Twenty-five years of research and application*. San Diego, CA: Edits Publishers, 1976.

16. Harburg, E., Erfurt, J. C., Hauenstein, L. S., Chape, C., Schull, W. J., & Schork, M. A. Socio-ecological stress, suppressed hostility, skin color, and black-white male blood pressure: Detroit. *Psychosomatic Medicine*, 1973, *35*, 276-296.

17. Hardyck, C. D., Chun, K., & Engel, B. T. Personality and marital-adjustment differences in essential hypertension. *Journal of Consulting Psychology*, 1966, *30*, 459.

18. Hardyck, C. D., Singer, M. T., & Harris, R. E. Transient changes in affect and blood pressure. *Archives of General Psychiatry*, 1962, *7*, 15-20.

19. Jacobson, E. *Progressive relaxation*. (2nd ed.) Chicago: University of Chicago Press, 1938.

20. Kidson, M. A. Personality factors in hypertension. *Australian and New Zealand Journal of Psychiatry*, 1971, *5*, 139-145.

21. Kirkendall, W. M., & Nottebohm, G. A. Essential hypertension. In J. Genest, E. Koiw, & O. Kuchel (Eds.), *Hypertension*. New York: McGraw-Hill, 1977.

22. Kirscht, J. P., & Rosenstock, I. M. Patients' problems in following recommendations of health experts. In G. C. Stone, F. Cohen, & N. E. Adler (Eds.), *Health psychology - A handbook*. San Francisco: Jossey-Bass, 1979.

23. Lewinsohn, P. M. Personality correlates of duodenal ulcer and other psychosomatic reactions. *Journal of Clinical Psychology*, 1956, *12*, 296-298.

24. Mann, A. H. Psychiatric morbidity and hostility in hypertension. *Psychological Medicine*, 1977, *7*, 653-659.

25. McClelland, D. C. Inhibited power motivation and high blood pressure in men. *Journal of Abnormal Psychology*, 1979, *88*, 182-190.

26. McGinn, N. F., Harburg, E., Julius, S., & McLeod, J. M. Psychological correlates of blood pressure. *Psychological Bulletin*, 1964, *61*, 209-219.

27. Miller, C., & Grim, C. Personality and emotional stress measurement on hypertensive patients with essential and secondary hypertension. *International Journal of Nursing Studies*, 1979, *16*, 85-93.

28. Mossey, J. M. *The psychosocial and behavioral consequences of blood pressure intervention*. Doctoral dissertation, University of North Carolina at Chapel Hill, 1975.

29. Ostfeld, A. M. What's the payoff in hypertension research? *Psychosomatic Medicine*, 1973, *35*, 50-57.

30. Pilowsky, I., Spalding, D., Shaw, J., & Korner, P. I. Hypertension and personality. *Psychosomatic Medicine*, 1973, *35*, 50-57.

31. Qualls, P. J., & Sheehan, P. W. Capacity for absorption and relaxation during electromyograph biofeedback and no-feedback conditions. *Journal of Abnormal Psychology*, 1979, *88*, 652-662.

32. Robinson, J. O. A study of neuroticism and casual arterial blood pressure. *British Journal of Social and Clinical Psychology*, 1962, *2*, 56-64.

33. Rosenstock, I. M. Why people use health services. *Milbank Memorial Fund Quarterly*, 1966, *44*, 94-124.

34. Rosenstock, I. M. The health belief model and preventive health behavior. *Health Education Monograph*, 1974, *2*, 354-386.

35. Rose, R. N., & Levin, M. A. The crisis in stress research: A critical reappraisal of the role of stress in hypertension, gastrointestinal illness and female reproduction dysfunction. Session I: The role of stress in hypertension. *Journal of Human Stress*, 1979, *5*, 7-26.

36. Sackett, D. L., & Haynes, R. B. *Compliance with theraputic regimens*. Baltimore: Johns Hopkins University Press, 1976.

37. Sainsbury, P. Neuroticism and hypertension in an outpatient population. *Journal of Psychosomatic Research*, 1960, *12*, 261-273.

38. Sandberg, B., & Bliding, A. Problems and symptoms in army basic trainees with stress-induced hypertensive reactions. *Journal of Psychosomatic Research*, 1976, *20*, 51-59.

39. Schori, T. R., & Thomas, C. B. Precursors of premature disease and death: Rorschach and figure-drawing factors. *Psychological Reports*, 1977, *40*, 1115-1122.

40. Shapiro, A. P. An experimental study of comparative response of blood pressure to different noxious stimuli. *Journal of Chronic Diseases*, 1961, *13*, 293-311.

41. Spielberger, C. D. Anxiety as an emotional state. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Volume 1). New York: Academic Press, 1972.

42. Steward, I. M., & Lond, M. D. Headache and hypertension. *Lancet*, 1953, *1*, 1261.

43. Stradoudakis, J. P., Hamesma, R. J., & Russell, R. K. Psychological dynamics and self-perceptions of persons with essential hypertension. *Perceptual and Motor Skills*, 1976, *42*, 360-362.

44. Tellegen, A., & Atkinson, G. Openess to absorbing and self-altering experiences ("Absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, 1974, *83*, 268-277.

45. Tessler, R., & Mechanic, D. Psychological distress and perceived health status. *Journal of Health and Social Behavior*, 1978, *19*, 254-262.

46. Thomas, C. B. Precursors of premature disease and death. *Annals of Internal Medicine*, 1976, *85*, 653-68.

47. Weiner, H. Personality factors and the importance of emotional stress in hypertension. In J. Genest, E. Koiw, & O. Kuchel (Eds.), *Hypertension*. New York: McGraw-Hill, 1977.

48. Wood, W. G., Elias, M. F., Schultz, N. R., & Pentz, C. A., III. Hypertension and symptoms reported on the Cornell Medical Index. *Experimental Aging Research*, 1978, *4*, 421-431.

49. Wood, W. G., Elias, M. F., Schultz, N. R., & Pentz, C. A. III. Anxiety and depression in young and middle-aged hypertensive and normotensive subjects. *Experimental Aging Research*, 1979, *5*, 15-30.

Journal of Personality and Social Psychology
1980, Vol. 38, No. 5, 793-800

Enduring Dispositions in Adult Males

Paul T. Costa, Jr., Robert R. McCrae, and David Arenberg
Gerontology Research Center, National Institute on Aging, National Institutes of Health
Baltimore, Maryland

Retest coefficients for temperamental traits measured by the Guilford-Zimmerman Temperament Survey were assessed at 6- and 12-year intervals to determine the degree of stability in personality and to evaluate the hypotheses that (a) younger men will show lower stability than older men, and (b) traits related to neuroticism will be less stable than traits related to extraversion. Subjects were 460 male volunteer participants in the Baltimore Longitudinal Study of Aging, ranging in age from 17 to 85 at the time of first testing. Results showed uncorrected stability coefficients ranging from .59 to .87. No consistent evidence of lower stability in younger subjects was found, and neurotic and extraverted traits appeared comparably stable when corrected for unreliability. The replicated pattern of consistent stability across age groups and across traits is discussed in terms of its implications for the further study of aging and personality.

Many personality theories take a clear stand on the issues of constancy or change in personality in adulthood. Temperamental theories (e.g., Buss & Plomin, 1975) that postulate genetic influences and psychoanalytic theories that posit childhood determinants of personality would both predict stability in adulthood. Social learning theory (Bandura, 1977), which sees the determinants of personality in the changeable environment, and growth theories (e.g., Gould, 1978), which portray continuing inner development throughout the life span, would easily accommodate evidence of change. Questions regarding constancy or change in personality are empirical, and one purpose of this article is to address the issue with longitudinal data.

From a more pragmatic point of view, the stability of measured characteristics is a crucial question for those who wish to make long-term predictions of behavior or adaptation. These practical considerations have inspired research on the stability of occupational interests and values (Kelly, 1955; Strong, 1955) that until recently provided the major evidence on the topic. More recently, as part of their systematic attempts to describe changes in the aging individual, gerontologists (Botwinick, 1973; Neugarten, 1977) have been concerned with stability, growth, and change in personality characteristics over the life span.

Clinical evidence has often seemed to support the positions of both stability and change. Jung's clinical practice, Buhler's (1935) analysis of personal biographies, Levinson's (1978) intensive interviews, and Gutmann's (1964) projective tests have formed the basis for models of personality change with age. On the other hand, clinical evidence favoring stability is also to be found in the form of retrospective interviews (Reichard, Livson, & Peterson, 1962), and longitudinal studies using projective techniques (Britton & Britton, 1972; Skolnick, 1966) and Q-sorts (Block, 1971).

A few years ago, evidence on personality stability was so weak that a thorough and balanced review (Neugarten, 1964a) citing stability coefficients "of the order of .30 to .40" could reasonably conclude that "the implication is that there is at least as much change as there is stability" (p. 188). How-

---

Requests for reprints should be sent to Paul T. Costa, Jr., Chief, Section on Stress and Coping, Gerontology Research Center, National Institute on Aging, NIH, Baltimore City Hospitals, Baltimore, Maryland 21224.

ever, recent longitudinal studies employing objective personality measures (Costa & McCrae, 1977, 1978; Douglas & Arenberg, 1978; Siegler, George, & Okun, 1979) have begun to accumulate evidence in support of the hypothesis of predominant stability or constancy of personality in adulthood. The present study adds to the body of evidence by examining stability of individual rank order in the scales of the Guilford-Zimmerman Temperament Survey (GZTS).

Stability of Mean Level

Two rather different meanings are typically attached to the phrase *stability*. On the one hand, stability can mean the preservation of rank order in individual differences in some trait or characteristic. Generally expressed as a correlation coefficient, this kind of stability is independent of the absolute level of the trait and thus of any developmental changes that uniformly affect all the subjects.

On the other hand (and perhaps more commonly), stability can be understood to mean a constancy of level. As such it contrasts with *change in mean level*, which usually refers to a systematic increase or decrease in a variable, perhaps attributable to maturation or to some treatment such as psychotherapy.

Since maturation is not an experimentally manipulable variable, studies on developmental changes in adult personality have had to rely on multivariate designs that eliminate the most plausible rival hypotheses to maturation. Cross-sectional studies, although the most convenient and the most popular, confound maturation with generational differences. Simple test-retest longitudinal studies are susceptible to distortion due to the influence of time-of-measurement or practice effects. The most adequate designs proposed involve the addition of cross-sequential and time-sequential analyses (Baltes, 1968; Schaie, 1965), which provide a more comprehensive identification of change phenomena (Baltes, Reese, & Nesselroade, 1977). Unfortunately, they are expensive and time consuming and have only rarely been employed.

Recent analyses of stability of mean levels (Costa & McCrae, 1978; Douglas & Arenberg, 1978; Siegler, George, & Okun, 1979) have found little evidence of maturational change in mean levels. The Douglas and Arenberg (1978) study is of particular importance, since it included cross-sequential and time-sequential designs in analyzing a large sample on a standard personality instrument over a considerable span of time. Only two variables showed effects attributable to maturation—masculinity and general activity declined—and these effects were quite modest in magnitude, amounting to about one-eighth of one standard deviation change over a 6-year period.

Longitudinal Stability of Relative Ordering

Questions of maturational change in the level of a trait are meaningful only if an enduring trait is under investigation. Traits are often *defined* as enduring dispositions, and this definition has led to some confusion about the meaning of stability in measured traits. Many researchers simply assume that any measured characteristic bearing the label "trait" must "by definition" endure over time. Within such an interpretation, empirical evidence of stability is often viewed as vaguely tautological. Rather than being part of the definition, temporal stability of relative ordering might more properly be regarded as a criterion against which any proposed trait must be judged. It is incumbent on the trait theorist to demonstrate that dispositions do indeed endure over extended periods of time.

Although it is apparently not well known, a considerable literature already exists concerning the stability of individual differences in objectively measured traits (Block, 1977; Costa & McCrae, in press; Moss & Susman, in press). Correlations as high as .84 for substantial samples have been reported over a period of 10 years (Costa & McCrae, 1977). Although further evidence on this point should be welcome, enough has already been reviewed to suggest that more specific hypotheses should become the focus of interest. Two empirical generalizations in particular seem to emerge. Moss and Susman, in an extensive review of stability and change throughout the life span, argue that certain socially desirable traits that can be interpreted as falling into the broad domain of extraversion are particu-

Table 1
Mean Levels of GZTS Scales at Three Administrations

Scale	Administration			*SD*[a]
	First	Second	Third	
General activity	17.14	16.82	16.32	6.33
Restraint	19.68	19.26	19.56	4.27
Ascendance	15.49	15.54	15.82	5.56
Sociability	18.59	18.74	18.98	6.34
Emotional stability	21.89	21.83	21.94	4.94
Objectivity	21.60	21.61	21.60	4.42
Friendliness	18.72	18.00	17.81	5.16
Thoughtfulness	18.30	17.54	17.73	4.72
Personal relations	22.98	22.21	21.75	4.15
Masculinity	20.75	20.48	20.04	3.52

Note. Means for 114 subjects with complete data at all three administrations. GZTS = Guilford Zimmerman Temperament Survey.

[a] Standard deviation at first administration.

larly stable and that change is more characteristic of undesirable traits, which might be construed as elements of neuroticism. In the present study, the GZTS scales of general activity, sociability, and ascendance are taken as representative of the social activity or extraversion domain, whereas low emotional stability, objectivity, friendliness, and personal relations represent the emotional health or neuroticism domain. These groupings are based on numerous factor analyses reported in the GZTS handbook (Guilford, Zimmerman, Guilford, 1976).

Moss and Susman also point out that in studies of children and young adults, stability coefficients tend to increase as older age groups are studied. Gerontological concepts of rigidity (Chown, 1961) and interiority (Neugarten, 1964b) also suggest that personality should show increasing stability with age. The present article will examine the 6- and 12-year stability of the GZTS scales, with particular attention to the question of whether (a) older cohorts show greater stability than younger cohorts and (b) scales measuring degrees of introversion–extraversion show higher stability than scales of adjustment–neuroticism.

Method

Subjects

Participants in the Baltimore Longitudinal Study are a highly select group of male volunteers, predominantly white, who agree to return for testing at fixed intervals. The vast majority have earned at least a college degree, work in (or are retired from) scientific, professional, or managerial positions, and are in good health (Stone & Norris, 1966). Data in this article come from responses of participants who entered the study from late 1958 through 1978. Douglas and Arenberg (1978) provide data on sample attrition.

Procedure

Each subject was given the standard GZTS instructions individually and completed the questionnaire during the remainder of his 3-day visit to the Gerontology Research Center. For each item, subjects choose *Yes*, *No*, or *?*. Each scale consists of 30 items, but only yes and no responses contribute to the scale score. A measure was invalidated for any scale with more than three *?* responses—a procedure suggested by Guilford and Zimmerman (1949). Therefore, small variations in the number of subjects will be seen on stability coefficients for different scales.

Subjects were readministered the GZTS approximately every 6 years. Because of complications in scheduling, a few subjects took the test 2 years in succession or failed to take the first retest but did take the second. To maintain consistency of time interval and number of administrations, longitudinal analyses are limited to subjects who took their second GZTS 5.0 to 7.9 years after their first (M = 6.6 years, n = 460) and to those who took their third GZTS 11.0 to 15.4 years after the first (M = 12.9 years, n = 222).

Three age groups were formed: young (17–44 years, M age 36.7, n = 145); middle (45–59 years, M age 51.5, n = 183); and old (60-85, M age 67.9, n = 132).

Table 2
Six-Year Retest Coefficients for GZTS Scales in Different Age Groups

Scale	Total (17–85)	Age: Young (17–44)	Age: Middle (45–59)	Age: Old (60–85)
General activity	.83 (410)	.79 (128)[a]	.86 (166)	.83 (116)
Restraint	.71 (418)	.68 (130)	.69 (168)	.70 (120)
Ascendance	.82 (401)	.79 (126)	.84 (160)	.80 (115)
Sociability	.81 (393)	.85 (126)[b]	.83 (160)[c]	.68 (107)
Emotional stability	.74 (427)	.84 (137)[d]	.72 (173)	.77 (117)
Objectivity	.71 (405)	.69 (129)	.73 (160)	.70 (116)
Friendliness	.77 (406)	.78 (127)	.78 (164)	.75 (115)
Thoughtfulness	.72 (418)	.84 (132)[d, e]	.70 (167)	.75 (119)
Personal relations	.73 (385)	.71 (125)	.75 (161)	.73 (99)
Masculinity	.75 (417)	.77 (135)	.70 (164)	.76 (118)
M stability	.77	.76	.77	.75

Note. *n*s are given in parentheses in columns; numbers in parentheses in column headings are age at first time. All correlations significant at $p < .001$. GZTS = Guilford-Zimmerman Temperament Survey.
[a] Difference between Young and Middle significant at $p < .05$. [b] Difference between Young and Old significant at $p < .01$. [c] Difference between Middle and Old significant at $p < .01$. [d] Difference between Young and Middle significant at $p < .01$. [e] Difference between Young and Old significant at $p < .05$.

Results

Table 1 gives the mean levels of the 10 GZTS scales for the subsample of 114 men who had valid data at all three administrations. Comparisons with handbook norms for college males show that the present sample is somewhat higher (up to one standard deviation) on the four scales measuring emotional health vs. neuroticism (emotional stability, objectivity, friendliness, and personal relations); no other differences are greater than one-quarter standard deviation. Level changes

Table 3
Twelve-Year Retest Coefficients for GZTS Scales in Different Age Groups

Scale	Total (20–76)	Age: Young (20–44)	Age: Middle (45–59)	Age: Old (60–76)
General activity	.77 (192)	.77 (60)	.82 (93)	.78 (39)
Restraint	.72 (193)	.61 (62)	.74 (94)	.76 (37)
Ascendance	.83 (194)	.85 (62)	.85 (95)	.77 (37)
Sociability	.74 (182)	.64 (62)[a]	.81 (88)	.66 (32)
Emotional stability	.70 (203)	.63 (68)	.76 (96)	.71 (39)
Objectivity	.69 (191)	.66 (64)	.76 (87)	.59 (40)
Friendliness	.74 (193)	.74 (64)[b]	.68 (88)[c]	.87 (41)
Thoughtfulness	.73 (199)	.78 (64)	.71 (94)	.71 (41)
Personal relations	.68 (188)	.70 (62)	.64 (89)	.73 (37)
Masculinity	.72 (200)	.73 (66)	.71 (94)	.70 (40)
M stability	.73	.72	.75	.73

Note. *n*s are given in parentheses; numbers in parentheses in column headings are age at first time. All correlations significant at $p < .001$. GZTS = Guilford-Zimmerman Temperament Survey.
[a] Difference between Young and Middle significant at $p < .05$. [b] Difference between Young and Old significant at $p < .05$. [c] Difference between Middle and Old significant at $p < .01$.

Table 4
Observed Retest Coefficients for Three Intervals and Estimated Reliability and Stability Coefficients for "True" Scores

Scale	r_{12}	r_{23}	r_{13}	Reliability	12-year stability
General activity	.83	.84	.80	.88	.92
Restraint	.75	.75	.71	.80	.89
Ascendance	.81	.85	.85	.82	1.00[a]
Sociability	.84	.82	.75	.91	.82
Emotional stability	.77	.83	.71	.89	.80
Objectivity	.77	.82	.74	.86	.86
Friendliness	.81	.78	.77	.83	.93
Thoughtfulness	.73	.76	.71	.78	.91
Personal relations	.70	.73	.68	.75	.91
Masculinity	.74	.77	.73	.79	.92

Note. Coefficients calculated with formulae of Heise (1969) for a subsample of 114 men with complete data at three times.
[a] Observed 12-year retest coefficient greater than estimated reliability.

in the present sample over time are small (cf. Douglas & Arenberg, 1978).

Stability of rank order of individual differences was examined through Pearson correlations for 6- and 12-year intervals. Stability coefficients were calculated for total sample and for three age groups, classified by their age at first administration.

Table 2 gives the six-year stability coefficients for GZTS scales between first and second administrations. Table 3 gives the 12-year coefficients between first and third administrations. A z transformation was used to compute the mean stability coefficients. All of these coefficients, which range from .59 to .87, are statistically significant ($p < .001$).

Under the hypothesis that stability should be greater in older age groups, one-tailed tests of the significance of differences between correlations were computed for each pair of age groups on each scale. Of the 30 comparisons, at each interval, six were significant for the 6-year interval, three for the 12-year interval. Of these nine significant differences, four were in the predicted direction, five in the opposite direction. None of the specific findings for scales at the 6-year interval are replicated at the 12-year interval.

Finally, Table 4 gives the estimated reliability and stability of the "true" scores (Heise, 1969) in a subsample of the 114 subjects of all ages who had complete data for all scales at the three times. Also in this table are simple retest coefficients for Times 1 to 2, 2 to 3, and 1 to 3. The similarity of correlations in the first and second 6-year periods is further evidence of the continuing stability of traits. It is also noteworthy that 6-year retest coefficients in this most select group are quite comparable to those presented in Table 1 for a more inclusive group of subjects who may have invalidated some scales by the overuse of the ? response or who may have dropped out of the study after the second administration of the GZTS.

Discussion

Stability of Rank Ordering

The retest coefficients presented here are among the highest in the literature for so long a period of time. Indeed, these would be noteworthy were they 2-week retest reliability coefficients instead of 12-year stability coefficients. Block (1977) advocates the use of statistical correction for attenuation due to unreliability of measurement, and Heise (1969) has developed from path-analytic considerations a formula for estimating the "true" stability of measures, which are presented in Table 4. Simple, uncorrected Pearson correlations, however, have much to recommend them, since they are straightforward and familiar, and the correlations seen here reach impressively large magnitudes despite error of measurement.

The question of increasing stability with age in adulthood seems to be clearly answered in the negative; 6- and 12-year stability coefficients are quite similar for three age cohorts whose members have an initial age range of nearly 70 years. Statistically significant differences in the magnitude of stability coefficients are scattered and inconsistent and do not support the hypothesis. Increased stability with age may be found among children and adolescents, but by young adulthood, stability in these dimensions of temperament is so high—near the limits of reliability of the instrument—that a ceiling effect diminishes the likelihood of any further increase in stability.

The question of differential stability for different traits is more difficult to answer, since change and error of measurement are confounded in retest coefficients. It can be seen that for the total sample, the three traits that constitute the extraverted factor of social activity—general activity, sociability, and ascendance—have mean coefficients of .82 and .78 for 6- and 12-year intervals, whereas the neurotic traits of low emotional stability, objectivity, friendliness, and personal relations show corresponding coefficients of .74 and .70. The latter are certainly lower, accounting for only about three-fourths as much variance. This result is comparable to a finding with the 16 PF (Costa & McCrae, 1977), in which 10-year coefficients ranged from .70 to .84 for extraversion and from .58 to .69 for anxiety or neuroticism. It may be the case that extraordinary stresses produce temporary neurotic tendencies in some people or that the distress attendant on neurotic traits leads some individuals to change, either by themselves or with the aid of friends or professional therapists.

On the other hand, it may also be the case that neuroticism is more difficult to measure reliably. The estimated stabilities of "true" scores, which correct for unreliability, are given in Table 4. These show an average value of .91 for the three extraversion scales and .88 for the four neuroticism scales, suggesting that there is little basis for inferring differential stability between these two domains of traits.

Comments and Conclusions

The data on which this article was based are taken from a single self-report instrument applied to a select male sample, and methodological artifacts—including sample selection and attrition, social desirability, and response sets—may have inflated the correlations. But a similar pattern of results in other samples using other objective measures (Costa & McCrae, 1977; Leon, Gillum, Gillum, & Gouze, 1979) as well as ratings (Block, 1971; Mussen, Eichorn, Honzik, Bieber, & Meredith, Note 1) argues that the present results cannot wholly be dismissed as method variance. Indeed, the theme of stability in personality is being heard with increasing frequency from a variety of sources. The conclusion seems to hold for women as well as men (Siegler, George, & Okun, 1979; Mussen et al., Note 1), for Germans as well as Americans (Grombach, 1976), and for adolescents as well as elders (Bachman, O'Malley, & Johnston, 1978).

Admittedly, many things do change with age. Physical, perceptual, and cognitive abilities generally decline. Sexual vigor and interest decrease. Social roles alter drastically, and events like retirement and widowhood produce enormous changes in the behavior of individuals. Bromley (1978) has argued that changes in mobility and expressive capacity lead to radical changes in the "stimulus value" of old people and thus to change in *our perception* of their personality. And it can be admitted that there are unmistakable age changes in the specific behaviors that express enduring traits: activity in older persons is likely to be gardening rather than football. But all these changes do not amount to change in personality, for personality is not the totality of behavior but rather "a stable set of characteristics and tendencies that determine those commonalities and differences in the psychological behavior . . . that may not be easily understood as the sole result of the social and biological pressures of the moment" (Maddi, 1976, p. 9).

The search for changes in personality in adulthood does not seem to have been fruitful, but the personologist concerned with adult development and aging can profitably turn to other questions. The durability of dispositions despite biological, social, and cognitive changes must result from some form of adaptation: By what mechanisms are we enabled to assimilate the changing experiences of a lifetime to our own nature? How do we cope, adjust, adapt, or defend so as to preserve our essential characteristics unchanged in the face of all the vicissitudes and transitions of adulthood and old age? These have become crucial questions for the student of personality.

Reference Note

1. Mussen, P., Eichorn, D. H., Honzik, M. P., Bieber, S. L., & Meredith, W. M. *Continuity and change in women's characteristics over four decades.* Manuscript submitted for publication, 1979.

References

Bachman, J. G., O'Malley, P. M., & Johnston, J. *Adolescence to adulthood: Change and stability in the lives of young men.* Ann Arbor, Mich.: Institute for Social Research, 1978.

Baltes, P. B. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, 1968, *11*, 145–171.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. *Life-span developmental psychology: Introduction to research methods.* Monterey, Calif.: Brooks/Cole, 1977.

Bandura, A. *Social learning theory.* Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Block, J. *Lives through time.* Berkeley, Calif.: Bancroft Books, 1971.

Block, J. Advancing the psychology of personality: Paradigmatic shift or improving the quality of research. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology.* Hillsdale, N.J.: Erlbaum, 1977.

Botwinick, J. *Aging and behavior: A comprehensive integration of research findings.* New York: Springer, 1973.

Britton, J. H., & Britton, J. O. *Personality changes in aging: A longitudinal study of community residents.* New York: Springer, 1972.

Bromley, D. B. Approaches to the study of personality changes in adult life and old age. In A. D. Isaacs & F. Post (Eds.), *Studies in geriatric psychiatry.* New York: Wiley, 1978.

Buhler, C. The curve of life as studied in biographies. *Journal of Applied Psychology*, 1935, *19*, 405–409.

Buss, A. H., & Plomin, R. *A temperament theory of personality development.* New York: Wiley, 1975.

Chown, S. Age and the rigidities. *Journal of Gerontology*, 1961, *16*, 353–362.

Costa, P. T., Jr., & McCrae, R. R. Age differences in personality structure revisited: Studies in validity, stability, and change. *Aging and Human Development*, 1977, *8*, 261–275.

Costa, P. T., Jr., & McCrae, R. R. Objective personality assessment. In M. Storandt, I. C. Siegler, & M. F. Elias (Eds.), *The clinical psychology of aging.* New York: Plenum Press, 1978.

Costa, P. T., Jr., & McCrae, R. R. Still stable after all these years: Personality as a key to some issues in aging. In P. B. Baltes & O. G. Brim (Eds.), *Life span development and behavior* (Vol. 3). New York: Academic Press, in press.

Douglas, K., & Arenberg, D. Age changes, cohort differences, and cultural change on the Guilford-Zimmerman Temperament Survey. *Journal of Gerontology*, 1978, *33*, 737–747.

Gould, R. L. *Transformations.* New York: Simon & Schuster, 1978.

Grombach, H. H. Consistency and change of personality variables in late life. In H. Thomae, (Ed.), *Patterns of aging: Findings from the Bonn Longitudinal Study of Aging.* Basel, Switzerland: Karger, 1976.

Guilford, J. P., & Zimmerman, W. S. *The Guilford-Zimmerman Temperament Survey: Manual of instructions and interpretations.* Beverly Hills, Calif.: Sheridan Supply Co., 1949.

Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. *The Guilford-Zimmerman Temperament Survey handbook: Twenty-five years of research and application.* San Diego, Calif.: Knapp, 1976.

Gutmann, D. L. An exploration of ego configurations in middle and later life. In B. L. Neugarten (Ed.), *Personality in middle and later life.* New York: Atherton Press, 1964.

Heise, D. R. Separating reliability and stability in test-retest correlation. *American Sociological Review*, 1969, *34*, 93–101.

Kelly, E. L. Consistency of the adult personality. *American Psychologist*, 1955, *10*, 659–681.

Leon, G. R., Gillum, B., Gillum, R., & Gouze, M. Personality stability and change over a 30 year period—middle age to old age. *Journal of Consulting and Clinical Psychology*, 1979, *47*, 517–524.

Levinson, D. J., Darow, C. N., Klein, E. B., Levinson, M. H., & McKee, B. *The seasons of a man's life.* New York: Knopf, 1978.

Maddi, S. R. *Personality theories: A comparative analysis* (2nd ed.). Homewood, Ill.: Dorsey Press, 1976.

Moss, H. A., & Susman, E. J. Constancy and change in personality development. In O. G. Brim Jr., & J. Kagan (Eds.), *Constancy and change in human development.* Cambridge, Mass.: Harvard University Press, in press.

Neugarten, B. L. Personality change over the adult years. In J. E. Birren (Ed.), *Relations of development and aging.* Springfield, Ill.: Charles C Thomas, 1964. (a)

Neugarten, B. L. Summary and implications. In B. L. Neugarten (Ed.), *Personality in middle and late life.* New York: Atherton Press, 1964. (b)

Neugarten, B. L. Personality and aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold, 1977.

Reichard, S., Livson, F., & Peterson, P. G. *Aging and personality*. New York: Wiley, 1962.

Schaie, K. W. A general model for the study of developmental problems. *Psychological Bulletin*, 1965, *64*, 92–107.

Siegler, I. C., George, L. K., & Okun, M. A. Cross-sequential analysis of adult personality. *Developmental Psychology*, 1979, *15*, 350–351.

Skolnick, A. Stability and interrelationships of thematic test imagery over twenty years. *Child Development*, 1966, *37*, 389–396.

Stone, J. L., & Norris, A. H. Activities and attitudes of participants in the Baltimore Longitudinal Study. *Journal of Gerontology*, 1966, *21*, 575–580.

Strong, E. K., Jr. *Vocational interests 18 years after college*. Minneapolis: University of Minnesota, 1955.

Schaie, K.W. (Ed.),
Longitudinal Studies of Adult Psychological Development.
New York: Guilford, 1983

Recent Longitudinal Research on Personality and Aging

PAUL T. COSTA, JR., ROBERT R. McCRAE, AND
DAVID ARENBERG

Methods and Models in the Study of Personality and Aging

Alternative Approaches to Personality

Personality psychology as a discipline is more unified by history than by common methods, theories, or even goals. Psychoanalysis, social learning theory, and measurement-based trait psychology all share the pages of introductory texts on personality, but little else. In the more specialized field of personality and aging, this problem is magnified. The psychodynamic thinking of Jung and Erikson and, more recently, of Gould (1978) and Levinson (Levinson, Darrow, Klein, Levinson, & McKee, 1978) has represented the major theorizing in the area, but most research has been based on the use of trait measures, with little or no relationship to these conceptions.

The significance of these circumstances is that no fully integrated review of "personality and aging" is really possible. The reviewer must choose either an eclectic presentation of the diverse work in the field or an incomplete, but focused, synthesis of some specific approach to personality and aging. In reviewing our own longitudinal work, we must necessarily choose the latter. The reader, however, should bear in mind that this chapter does not pretend to be an exhaustive survey of the field. In particular, when we speak of the empirically demonstrated stability of

PAUL T. COSTA, JR., AND ROBERT R. McCRAE. Section on Stress and Coping, Laboratory of Behavioral Sciences, Gerontology Research Center, National Institute on Aging, National Institutes of Health, Baltimore, Maryland.

DAVID ARENBERG. Section on Learning and Problem Solving, Laboratory of Behavioral Sciences, Gerontology Research Center, National Institute on Aging, National Institutes of Health, Baltimore, Maryland.

personality, we are referring to only one definition of personality. We do not know if ego functions, instinctual impulses, cognitive constructs, approach–avoidance gradients, needs for achievement, or psychophysiological expressions of emotion change with age. Relatively little research has been done in these areas, and conclusions would be premature. We can, however, speak about the tradition in personality research that identifies personality with individual differences in dispositions, including interpersonal relations, emotional responsiveness, and receptivity to experience.

The choice of a measurement-based trait approach to the study of aging was an outgrowth of longitudinal research strategies that have been widely adopted. Large-scale interdisciplinary studies of aging that included personality variables adopted the convenient self-report trait inventories as a way of gathering standardized data on subjects. Interviews and projective methods, which might yield either trait content or process data, depending on the way they are scored and analyzed, have been used in some studies (Berkeley, Duke), but they require an extraordinary investment of time and effort. By contrast, self-report inventories with their objective scoring procedures eliminate both the necessity for clinical expertise in interpretation and the possibility of interpretative bias. The rapid, uniform quantification from such personality inventories permits rigorous assessment of reliability and validity of the measures, which can then provide valuable baseline data on the condition of the individual at some particular time. The standardization of these measures makes possible the meaningful readministration of the same measures at a later time, so that longitudinal comparisons are straightforward. Finally, most of the cross-sectional studies comparing adults of different ages have used self-report measures, and comparison of longitudinal with cross-sectional results is most readily handled within the trait perspective.

A Model of Personality Traits

Within the trait approach to the study of personality and aging, the major obstacle to an understanding of the literature is the lack of shared conceptual models. Most theories of personality do not define the elements of personality or specify the major variables to which attention should be given. Instead of a systematic enumeration of the elements of personality, most personologists content themselves with the elaboration of one or a few constructs that they feel are important. In the absence of a shared definition of personality, constructs simply proliferate. Anxiety (Spielberger, 1972), Sensation Seeking (Zuckerman, 1979), Dogmatism (Rokeach, 1960), Repression-Sensitization (Byrne, 1964), Ascendance (Allport & Allport, 1928), Absorption (Tellegen & Atkinson, 1974), Depression (Beck, 1972), Activity (Buss & Plomin, 1975), and Authoritarianism (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) are among the more widely recognized traits. Scores of additional trait measures could be enumerated, but to do so would only confuse the issue. Is each of these to be regarded as a separate dimension of personality? If so, then each would need to be studied separately, and conclusions about the relationships of aging and personality would be impossible. Probably none has been the subject of enough studies to warrant a review and conclusion.

But if it were possible to group these traits into a handful of larger domains, then separate studies using ostensibly different trait measures could be compared, and generalizations about the whole domain might be drawn. In his classic paper "Traits Revisited," Gordon Allport (1966) acknowledged "the powerful contributions of Thurstone, Guilford, Cattell and Eysenck, based on factor analysis . . . [which] should provide eventually a satisfactory taxonomy of personality and of its hierarchical structure" (p. 3). The factor-analytic personality tradition provides an empirical basis for this kind of organizational simplification. Thus Anxiety, Depression, and Repression-Sensitization all covary within the domain of Neuroticism; Sensation Seeking, Ascendance, and Activity are classed as Extraversion; and Dogmatism, Absorption, and Authoritarianism are all elements of Openness to Experience. Our experience has convinced us of the utility of this three-domain model of personality. While some personality variables (such as locus of control) probably do not fall in any of our domains, a great many do. As a basis for organizing the literature on personality and aging, we find this model invaluable and rely on it explicitly in this review.

The bulk of research summarized in this chapter is also tied to self-report procedures, which have always been peculiarly associated with trait systems. However, the conceptualization of traits and their organization into domains is not a function of the method of assessment used. Our current program of research in the Baltimore Longitudinal Study of Aging employs performance tests of cognitive style; projective methods, including the Thematic Apperception Test (TAT) and the Holtzman Inkblot Test; semistructured interviews with *Q*-sort ratings; and spouse ratings. The data from these sources are conceptualized within the same three-domain model, but may also allow the application of different approaches to personality and aging in future years.

Most of the uniquely longitudinal research that we have conducted has dealt with the question of stability or change in personality, and a large section of this chapter is devoted to a discussion of the findings on this issue. Since all this research employs self-report methods, a consideration of their limitations is in order, and some new longitudinal analyses on response sets are presented in the subsequent section. But the real utility of personality research is in explaining phenomena outside of personality proper, and thus the last major section highlights some of the unique contributions of longitudinal research to an understanding of the relationships between personality and other variables such as well-being and perceptions of health. Finally, we indicate some of the new directions for longitudinal research to which a stable trait model of personality leads.

The Research Context

Much of the research described in this chapter has been conducted as a part of the Baltimore Longitudinal Study of Aging. Personality research represents only a small part of the overall Baltimore study program, which is an intensive, interdisciplinary study of the aging process. Since 1958, volunteers have been seen every 1 to 2 years, with regular medical examinations and periodic cognitive and psychological testing. Subsets of subjects have participated in a variety of special studies, and the data from all these sources can be used to characterize subjects.

Until 1977, the Baltimore study looked at men only, and thus longitudinal data are available only for them. Beginning with an initial pool of scientists who volunteered for the study, subjects have been continuously recruited, usually by friends or relatives who were already participants. New subjects are accepted from a waiting list of volunteers, and an attempt to maintain roughly even numbers of subjects in each age decade from the 20s to the 80s has been the primary basis for selection from the list. For the intensified study of personality, which began in 1978 with the creation of a section on stress and coping, additional subjects were needed. The spouses of volunteers were invited to participate in a stress-and-coping project, which involved only responses to questionnaires by mail. In this way, the number of women available for study was substantially increased.

A number of distinctive features of the Baltimore study result from these sampling procedures. Because subjects entered the study continuously over a 20-year period, it is possible to conduct sequential analyses on independent samples for many of the measures. The sample itself is also distinctive. Participants in the study are a community-dwelling, generally healthy group of volunteers who are committed to the research goals of the program. Among the men, the majority (80%) work in or are retired from scientific, professional, or managerial positions. Almost all (93%) are high school graduates, and 71% are college graduates. The women who have joined the study and the wives of the male subjects who participate in the mail studies generally share these economic and educational advantages.

The Stability of Personality in Adulthood

Statistical Definitions of Stability or Change

The term "stability" has two different and largely independent meanings. A trait may be considered stable for a group if the mean level of the trait in a group of individuals is constant over time. This can occur if all individuals remain at the same level, or if increases offset decreases over the interval. Analysis of variance on repeated administrations is the usual way to determine if significant changes in mean level have occurred. However, stability can also be assessed in terms of a test–retest correlation coefficient. These coefficients will be higher if individuals maintain the same relative ordering on the trait over time, regardless of the level of the trait. If some developmental process leads to a *uniform* increase or decrease of a variable over time, it would have no effect on the retest coefficient.

The implication of these considerations is that, except in the artificial case in which all individuals score identically on repeated administrations of a test, the issues of mean-level stability and retest stability must be addressed separately.

There is one final way in which stability or change in personality might be seen across the life span. Personality is often conceptualized in terms of the relationship between discrete variables, and these variables might change with age. The pattern of intercorrelations among a group of traits might alter with maturation; although this kind of change is least familiar, it is logically prior to the other kinds of stability or change. From the point of view of construct validity, what a test measures is determined by what it is correlated with. If the correlates of a test change, then the test itself, or the construct it represents, has somehow changed. If this has happened, then problems of interpretation of retest coefficients or mean levels arise.

The most common method of comparing patterns of intercorrelations is by factoring the battery of tests and showing that the same (or different) factors emerge in different age groups or administrations. This method has been used more frequently in studying cognitive abilities than personality dispositions, but the logic is similar. "Age-comparative factor analysis" is a term occasionally used to designate this kind of analysis (Cunningham, 1978). Considering the relationships between traits as the "structure" of personality, we have usually referred to this kind of problem as one of "structural stability."

Differences or changes in personality-test factor structure would have methodological as well as theoretical implications for gerontologists. Theoretically, a large number of distinguishable factors might be taken as a sign of personality differentiation, a hallmark of development. Different organizations of personality variables, or different "syndromes," may appear with age, as traits take on new meanings or new functions and significance. Methodologically, if major differences were found in the factor structure of standard personality tests in older samples, the use and interpretation of the tests in these groups could be challenged. The major personality inventories, such as the Eysenck Personality Inventory (EPI), the Guilford-Zimmerman Temperament Survey (GZTS), and Cattell's 16 Personality Factors Questionnaire (16PF), relied explicitly on factor analysis in their development, and age invariance is required for maintaining their factorial validity across age. All tests must demonstrate their validity by consistent patterns of convergent and divergent relationships. In the extreme case, all previous test-development efforts (except that conducted on elderly samples) would be called into question in application to elderly populations, as would all research that has employed these tests in the study of aging.

Beginning with structural stability, we review in this chapter our findings, which point to constancy of structure, stability of mean levels, and consistency of individual differences across time. Together, these results lead to the conclusion that adult age *per se* has no noticeable effect on any of the domains of personality we have studied.

But let us make it quite clear that these findings do *not* prove that personality is unchangeable. Some individuals do change in one or more characteristics, for reasons not yet understood. It is reasonable to suppose that psychotherapeutic interventions can make real changes in personality, and a host of techniques, from cognitive-behavior modification to biochemical interventions, may have profound effects as yet undocumented. What we can say is that such changes, for better or worse, are not likely to happen to anyone simply as a result of growing older.

Age Invariance of Personality Structure

The Quest for Change in Cattell's 16 Personality Factors Questionnaire

Our first research efforts, in fact, concerned looking for structural differences in personality within different age groups. In 1975, cluster analysis appeared to be a "cutting-edge" methodological technique for examining personality structure, and we conducted cluster analyses of the 16PF scales within three age groups using the male subjects of the Veterans Administration's Normative Aging Study (NAS). The cross-sectional results first presented at an American Psychological Association (APA) symposium in 1975 were attended with great interest and even greater hope that, at last, using objective standardized measures and sophisticated statistics on a very large sample, a developmental lodestone had been uncovered. Instead, the results were actually the first formulation of an age-invariant, three-domain model of personality.

Cluster analyses of 16PF scales were conducted within each of three age groups: 140 men aged 25–34 ($M = 32$), the young group; 711 men aged 35–54 ($M = 44$), the middle-aged group; and 118 men aged 55–82 ($M = 60$), the old group. Three clusters were found for each age, accounting for about 21%, 14%, and 6%, respectively, of the total variance in each group (Costa & McCrae, 1976).

The first cluster, which accounted for nearly half the common variance in all age groups, contained Scales (low) C (Stable), Q_4 (Tense), O (Guilt-prone), L (Suspicious), and (low) Q_3 (Controlled). Originally we labeled it an Anxiety–Adjustment cluster, following Cattell's convention, although we now prefer to interpret the first cluster as representing Neuroticism. The composition of the Neuroticism cluster was identical or invariant in the three age groups. Similarly, the second cluster, containing Scales A (Outgoing), F (Happy-go-lucky), H (Adventurous), and Q_2 (Group-dependent), was constant across all three age groups. What captured our imagination were the differences in the third and smallest cluster. The third cluster in the young and middle-aged groups had two scales as its elements, with Scale M (Practical vs. Imaginative) common to both. The second element in the young group's cluster was Scale I (Tough-minded vs. Tender-minded). In the middle-aged group, Scale I was "replaced" by Scale Q_1 (Conservative, Respecting Established Ideas vs. Experimenting, Free-Thinking). In the old group, the third cluster contained Scale B (Bright) as well as Scales I, M, and Q_1.

If we examine momentarily the details of these age differences in the third cluster, the reader may be able to appreciate how we were swept along by the apparent "lawful changes" (even though they were only differences) in experiential phenomena. We interpreted the young group's third cluster as an openness to feelings and aesthetic sensitivity. The adjectival descriptions for high scorers on Scale I include "sensitive" and "intuitive"; for Scale M, they include "unconventional" and "imaginative." It seemed reasonable to infer that the combination of high I and high M

represented openness to feelings, whereas low *I* and low *M* ("unsentimental, logical, practical, narrow interests") represented affective closedness. The high pole of this cluster dimension seemed to characterize the idealism and romanticism of youth.

In the middle-aged group, *M* again appeared, but this time in conjunction with Q_1 (Liberal Thinking). For these men, openness seemed to appear more in the realm of ideas and values than feelings. We speculated that the familial and professional obligations of the middle-aged man had transformed his concern from impractical feelings to more consequential ideas. Finally, the third cluster in the old group suggested even more tempting interpretations. The feelings cluster and the ideas cluster were merged, along with Scale *B*, which measures intelligence. This appeared to be the marvelous developmental synthesis of opposing psychological functions that C. G. Jung (1933) had promised us, and perhaps this synthesis was the basis of wisdom, at least for those old men who remained open to experience. Indeed, the integration of these processes might well be the end point of psychological development. The differences in cluster structure were certainly consistent with such an explanation.

We attempted to rule out the possibility that the cluster structure differences we observed were statistical artifacts of the particular age groups we constructed and were able to replicate roughly the third cluster differences in two repartitionings of the sample. But we recognized and stated in the 1976 article an even more important limitation and caution: The data were cross-sectional and did not provide direct evidence of structural changes in individuals.

Needless to say, we were greatly encouraged by the publication of these findings and by the enthusiastic reception they received from many of our colleagues. However, we felt an urgent need to document longitudinally these important changes in personality organization and structure. In 1975, we had administered the 1967 edition of the 16PF as a longitudinal retest, but major changes in the item composition of the scales had occurred since the 1962 edition, making direct longitudinal comparisons impossible. We therefore decided to readminister Scales *I*, *M*, and Q_1 from the 1962 edition, Form A, in 1977. Our original cross-sectional data had been based on combination A and B forms, so it was necessary to separate A from B scales in the original data. An examination of the new data showed no evidence of the cluster changes we had hypothesized and hoped for, and, to make matters worse, the clusters could not be located in the original data when only the A form was analyzed. Only a simple I-and-M cluster could be identified at any age.

It turned out to be the case that the variations in the original cluster analyses on combined A and B forms were largely the result of error due to the unreliability of the 16PF scales. Our conclusion was that the age-specific clusters should not be regarded as successive phases in a developmental sequence, but as a series of crude approximations to an underlying age-invariant dimension of Openness to Experience.

We published a letter to the editor of the *Journal of Gerontology* (Costa & McCrae, 1978a) to inform readers of that journal of our results, as well as a full report in another journal (Costa & McCrae, 1977), and we have cited these corrective findings in several other articles. Yet even several years later, the original finding is still discussed, while the failures to replicate are ignored (e.g., Thomae, 1980). Bad news apparently travels slowly.

The good news to us was that a third dimension, Openness to Experience, had been identified, and subsequent research using more reliable measures of openness to fantasy, aesthetics, feelings, actions, ideas, and values continued to form a third dimension of personality alongside the ever-present Extraversion and Neuroticism. Subsequent cross-sectional analyses (Costa & McCrae, 1980c) provided clear evidence that the composition of this domain of personality was invariant across the adult age range.

Constancy of Structure in the Guilford-Zimmerman Temperament Survey

It was, therefore, with an expectation of invariance that we approached the personality data of the Baltimore Longitudinal Study of Aging, where the GZTS had been given to subjects over a period of 20 years. In addition to cross-sectional comparisons of the factor structure of the GZTS scales in three different age groups, we performed longitudinal and time-of-measurement analyses as well (McCrae, Costa, & Arenberg, 1980). Data were obtained from men who entered the study from late 1958 through 1978). The age range of the 769 men at the first time of administration was from 17 to 97, with a mean of 50 years. Second-administration data (5.0 to 7.9 years later) were obtained from 346 men aged 26-91 ($M = 57.6$ years); third-administration data (11.0 to 15.4 years after first administration) were from 171 men aged 33-86 ($M = 62.0$ years). To assess possible structural differences stemming from time-of-measurement effects, the sample was divided into two groups: 455 men who completed the GZTS before July 1968 (age range = 17-83, $M = 52.1$ years) and 314 men who first completed the GZTS after that date (age range = 18-96, $M = 45.6$ years). To avoid possible confounding with practice effects, these last analyses were limited to first-administration data. All factor analyses were restricted to subjects with valid scores on all ten GZTS scales.

Both principal components and principal axes factor analyses were examined. Three factors had eigenvalues greater than unity and accounted for similar amounts of variance in all eight analyses: 28.3% to 30.3% for the first factor; 20.8% to 22.6% for the second; and 11.8% to 13.7% for the third. After Varimax rotation, comparison of the two methods of factoring showed highly similar results. Only the principal components solutions are presented because these results are somewhat clearer.

Table 7.1 shows factor loadings across analyses of data from three administrations, three age groups, and two times of measurement. Emotional Stability, Objectivity, Friendliness, Personal Relations, and Masculinity are consistent definers of this Emotional Health versus Neuroticism factor; low Thoughtfulness is marginal. The post-1968 analysis shows small contributions from Restraint and Sociability. The inclusion of low Masculinity among the definers of Neuroticism is somewhat unusual; it may result either from the use of an exclusively male sample or from the fact that 10 of the 30 items in this scale concern susceptibility to the emotions of fear and disgust.

The pattern of General Activity, Ascendance, and Sociability seen in the second factor in all eight analyses has been labeled "Social Activity" by Guilford, but could also be identified as Social Extraversion. Emotional Stability, low Restraint, and occasionally low Friendliness show small contributions to this factor.

The third factor has been designated "Thinking Introversion" by Guilford, but it is not to be confused with the Introversion-Extraversion factors of Eysenck (1960) or Cattell (1973). Clearly composed of Restraint and Thoughtfulness across all eight analyses, it shows a small contribution from low Masculinity in some cases. On psychological grounds, the meaningfulness of the third factor is questionable. Typically, other personality tests do not yield such a factor, nor do Guilford, Zimmerman, and Guilford (1976) discuss any clinical significance or counseling application of the factor. As a measure of Sensation Seeking or Impulsivity, Restraint may more properly belong in the domain of Extraversion, and small, but consistent, negative loadings of Restraint on the second factor are consistent with this hypothesis.

Maturational changes in personality structure should appear in both longitudinal and cross-sectional comparisons. As is clear from Table 7.1, the major definers are the same at each time and in each age group. Small variations in loadings do occur, but they do not show a clear direction or pattern replicated across longitudinal and cross-sectional analyses.

To quantify these impressions of invariance, coefficients of factor

TABLE 7.1. Factor Loadings for GZTS Scales across Administrations, Age Groups, and Times of Measurement

	Longitudinal administrations			Cross-sectional age groups			Times of measurement	
	1st (*n* = 769)	2nd (*n* = 346)	3rd (*n* = 171)	17 to 45 (*n* = 314)	46 to 59 (*n* = 242)	60 to 97 (*n* = 213)	Pre-1968 (*n* =455)	Post-1968 (*n* = 314)
Factor I: Emotional Health versus Neuroticism								
General Activity	−12	−14	−03	−15	−06	−16	−10	−16
Restraint	28	18	16	22	27	23	23	33
Ascendance	07	09	13	07	09	−01	06	13
Sociability	22	17	12	22	16	21	12	32
Emotional Stability	73	73	72	67	76	73	76	69
Objectivity	86	84	85	85	87	87	87	84
Friendliness	77	76	72	78	76	76	74	79
Thoughtfulness	−25	−31	−36	−33	−24	−34	−30	−19
Personal Relations	73	71	67	72	71	72	74	71
Masculinity	46	50	57	56	48	49	46	44
Factor II: Social Activity or Extraversion								
General Activity	68	68	76	70	66	63	66	72
Restraint	−33	−36	−34	−19	−32	−37	−31	−34

Ascendance	86	88	86	85	87	84	86	87
Sociability	80	78	77	81	82	77	80	79
Emotional Stability	30	30	38	40	27	31	31	28
Objectivity	17	25	24	23	19	14	17	17
Friendliness	−30	−31	−36	−24	−28	−23	−31	−28
Thoughtfulness	21	19	23	16	11	14	16	07
Personal Relations	04	−01	03	13	06	−01	−01	14
Masculinity	−08	−05	00	−11	−10	−12	−09	−06
Factor III: "Thinking Introversion"								
General Activity	−10	−09	02	−11	−03	−15	−03	−17
Restraint	71	76	79	79	69	73	76	67
Ascendance	07	05	07	18	−04	07	05	10
Sociability	−04	−06	−31	−09	−02	−05	−10	−02
Emotional Stability	−13	−10	−09	−17	−12	04	−01	−24
Objectivity	−12	−07	−10	−09	−07	−03	−04	−22
Friendliness	15	12	−07	13	23	11	14	12
Thoughtfulness	81	80	74	79	78	82	78	84
Personal Relations	08	03	−17	15	03	−07	00	11
Masculinity	−33	−10	17	−13	−32	−12	−21	−44

Note. Varimax-rotated principal components. Decimal points omitted.

congruence (Gorsuch, 1974) were calculated between corresponding factors for administrations, age groups, and times of measurement. Used with principal components, as they are here, these coefficients are equivalent to the product-moment correlations between factor scores. All coefficients are above .98 for the first two factors; for the third and smallest factor, they range from .83 to .98.

Additional analyses were conducted on data from the second administration for subjects aged 25-45 ($n = 60$), 46-59 ($n = 154$), and 60-91 ($n = 132$), and on data from the third administration for subjects aged 32-62 ($n = 84$) and 63-86 ($n = 87$). In all five analyses, three factors had eigenvalues above 1.0, and, despite small sample sizes, generally similar structures were observed. These are reflected in congruence coefficients (when compared with the full first administration solution) ranging from .96 to .99 for the Neuroticism factor, .91 to .99 for the Extraversion factor, and .58 to .99 for the "Thinking Introversion" factor.

Finally, in order to parallel traditional longitudinal designs in which the same subjects are tested on successive occasions (mean intervals = 6.6 and 12.9 years), comparisons were made between different administrations, restricting subjects to those with complete data on all administrations. Although the number of subjects meeting this qualification at all three administrations was relatively small ($n = 123$), the same general pattern of definers was replicated at each administration. The congruence coefficients were .99, .98, and .83 for the first, second, and third factors, respectively. When analyses were conducted for the 324 subjects who had complete data on both first and second administrations, coefficients of congruence for the corresponding factors across the two administrations were .98, .99, and .98.

The age-invariance of factor structure in the GZTS in the sample cited was clearly evident. Definers of factors stand out in each case from the marginal elements that show slight variations from one analysis to another. Despite aging, attrition, and possible practice effects, the same pattern is seen at each administration. The only divergence from high factor congruence is found in the case of the third factor at the third administration, when a coefficient of .83 is observed. Although this slight structural variation could result somehow from repeated exposure to the test, it is interpreted more simply as the result of error in the smallest factor and smallest sample. In the data presented here, no meaningful difference can be seen when comparing measurements before and after 1968, yet the decades of the 1960s and 1970s are surely different enough to make a difference if the structure of the GZTS and similar personality tests were particularly sensitive to historical and cultural shifts.

Stability of Mean Levels

Cross-Sectional Studies: Minnesota Multiphasic Personality Inventory

Until recently, most of our knowledge concerning the descriptions of adult personality has come from one-time administration of various personality questionnaires to a variety of adult samples. Much of the literature on assessment of age differences in adult personality has compared nonmatched extreme groups of old and young on measures that often lack reliability or validity. The Minnesota Multiphasic Personality Inventory (MMPI), which has more desirable psychometric properties, is less relevant to normal personality-trait descriptions than to mental health or psychopathology. But since it is a widely used instrument, we shall briefly and critically review the literature on age differences in the MMPI.

Lawton, Whelihan, and Belsky (1980) review 11 studies that compare MMPI clinical scale score ranks of elderly to younger subjects. Seven of the 11 studies contain either older psychiatric patients, institutional residents, or medical patients. One study contains job applicants over the age range of 19–56 years; the remaining three studies include older community residents. Lawton *et al.* assert that depression is clearly elevated among the elderly along with hypochondriasis, while the "acting out" scales (*Pd* and *Ma*) are clearly lower. Scores of the elderly on the "psychotic triad" scales (*Pa, Pt, S*) were either similar to or lower than scores of younger people, with minor exceptions only in the Newcastle-upon-Tyne community sample.

In his excellent review of age and the MMPI, Gynther (1980) arrives at broadly similar conclusions to the effect that certain pathological features decline with age. From youth to young adulthood, there are declines in admission of rebelliousness towards authority, suspiciousness, autistic thinking, and impulsivity. Since scale ranks are ipsative, the apparent increase in certain neurotic features, particularly depression, is interpreted as a consequence of age-linked *decreases* in the scores for Scales 4 (*Pd*), 6 (*Pa*), 8 (*Sc*), and 9 (*Ma*). Along with dysphoric affect, health problems are said to be salient for the elderly, and health and happiness seem to emerge as characteristic personality problems for the elderly. Yet these results are not beyond challenge.

Zemore and Eames (1979) reported a most instructive set of findings that question the widely held belief that the aged are more depressed than any other group. They noted that studies based on symptom counts find greater evidence of depression over age 65 in contrast to studies based on clinical judgments, which find depressive disorder most frequently between the ages of 25 and 65. They hypothesized that symptom checklists for depression invariably include a variety of somatic complaints more likely to reflect declining physical health than depression. They argued that clinicians would not diagnose as depressive elderly who have few cognitive or affective symptoms of depression, but who admit to fatigue, constipation, and sleep disturbances. In a simple and straightforward study, they compared psychic (cognitive and affective) versus somatic symptoms of depression on the Beck Depression Inventory (BDI) among 424 first-year psychology students, 48 elderly long-term residents of an old-age home, and 31 community-residing elderly awaiting entrance to an old-age home. The results supported their hypothesis: Namely, the BDI scores of the elderly were significantly higher than those of the students only when the 7 somatic items were included, that is, using all 21 items. When only the 14 psychic item scores were compared, the mean scores (4.47 vs. 4.43) were virtually identical. Thus even institutionalized elderly and those awaiting entrance to institutions for the elderly did not show any more psychic symptoms of depression, although they did report more somatic complaints. The latter are likely to covary with the physiological changes that accompany aging.

The Zemore and Eames results should serve to remind us that findings of age-associated differences need to be interpreted correctly. Even the apparent age-associated increases in physical complaints need to be scrutinized carefully. In our longitudinal studies of symptom reporting, we have not found generalized increases in somatic complaints due to aging (as the cross-sectional literature would suggest). Instead, specific age-related symptoms, namely, sensory, cardiovascular, and genitourinary problems, were seen to increase (Costa & McCrae, 1980b). Interestingly, Gynther (1980) reports similar findings from MMPI studies for nonpsychiatric patients.

Cross-Sectional Studies: Factor-Based Inventories

Eysenck's two-dimensional model of personality, operationalized successively in the Maudsley Personality Inventory, the EPI, and (with the addition of a third dimension, Psychoticism) the Eysenck Personality Questionnaire, has occasionally been the object of cross-sectional studies. Eysenck (e.g., Eysenck & Eysenck, 1975) has consistently reported that younger subjects are higher in both Neuroticism and Extraversion than are older subjects, including both men and women. Older subjects score higher on the Lie scale, which may indicate greater social desirability, but might also be interpreted as showing higher levels of socialization.

Some of these cross-sectional trends for the EPI are paralleled by the cross-sectional results of Sealy and Cattell (1965), who gave the 16PF to a large sample of men and women aged 16–70. They found a significant decrease on the *F* (Happy-go-lucky) scale, which is a facet of Extraversion; an increase on the *C* (Stable) scale, which suggests lowered Neuroticism; and increases on the *G* (Conscientious) scale, showing higher Superego Strength or Socialization for the older subjects. The pattern of scale scores indicated statistically significant, but small, differences on the second-order factors of Anxiety and Extraversion.

Schaie (1959) found similar, but later occurring, introversion with his Social Responsibility scale. Fozard and Thomas (1975) observed that the decrease in Scale *F* (Happy-go-lucky) toward a more sober, serious, glum disposition is the most consistently observed age difference with the 16PF. Other studies have reported decreases in Scales *A* (Outgoing), *H* (Adventurous), and *E* (Assertive). These are recognized as signifying shifts away from extraversion, toward introversion. The first-order scales making up the Neuroticism domain from the 16PF (Scales *C, O, L,* Q_3, and Q_4) have shown inconsistent trends from one study to another. From Lawton *et al.*'s (1980) review of six cross-sectional studies using the 16PF, only Scale *G* (Conscientious) shows higher scores for older subjects across all studies.

The last group of cross-sectional studies discussed here employed the GZTS (Guilford *et al.*, 1976), showing generally similar findings. Bendig (1960) examined the GZTS mean score differences of men in four age groups, finding significantly lower scores on General Activity, Ascendance, Sociability, and Masculinity. Significantly elevated scores were observed for Restraint and Personal Relations. Wagner (1960) compared 150 male executives 45 years and older with 150 who were 35 years or younger, with results consistent with those of Bendig (lower Ascendance, lower Sociability, and higher Restraint scores for older executives). Unlike Bendig, Wagner also found significantly lower scores for the Emotional Stability and Objectivity scales, which are components of Emotional Health, as Guilford *et al.* (1976) call this second-order factor.

Cross-sectional findings for the ten GZTS scale on a very large sample of men covering a wide age range have been reported by Douglas and Arenberg (1978) on the Baltimore study participants. Five of the ten scales were significantly related to age in two subsamples defined in terms of the date of first administration of the GZTS: Sample A, 605 men (aged 17–98) tested prior to July 1968; and Sample B, 310 men tested between July 1968 and June 1974. Like both Wagner and Bendig, Douglas and Arenberg found Ascendance to be negatively correlated with age. General

Activity and Masculinity were also negatively correlated with age, as only Bendig previously showed. The other component of the Social Activity or Extraversion factor, Sociability, was not significantly correlated with age in both samples. Restraint and Friendliness scale scores were positively correlated with age ($r = .28$ and .17, respectively). The General Activity decade means decreased consistently after the 30s from 18.78 to 16.92 for the 60s, while mean scores of 14.79 and 13.46 were observed for the 70- and 80-year-old groups, respectively. Means for Ascendance decreased monotonically from 18.35 (20s) to 14.35 (80s) for Sample A (pre-1968), but less consistently in Sample B (post-1968). The differences in Masculinity were not apparent in either sample until age 60. As for the increases with age, Restraint means increased monotonically in Sample A from 17.40 (20s) to 21.93 (70s), but less consistently in Sample B. Friendliness mean scores showed a similar pattern to the Restraint scores for Sample A; that is, the 20s group mean of 14.95 increased to 19.13 for the 70s group.

Consulting the GZTS *Handbook* (Guilford *et al.*, 1976), one finds that the most consistent finding with regard to GZTS scores and age is the decrease in Ascendance (Scale *A*) reported in 12 different samples. The second most consistent finding was an increase (cross-sectionally) in Restraint (*R*) scores observed in nine different samples. Seven samples demonstrated decreases in Sociability with age, while four samples showed General Activity decreases with age. Age differences for the other GZTS scale scores are inconsistent and "conflict markedly" (1976, p. 107).

Guilford *et al.* conclude that "there would seem to be little risk in using the same GZTS norms with any age group despite the consistency of the scale score differences with age, since the correlations and mean differences are small" (p. 105).

Repeated-Measures Analyses

Because all of the previous studies mentioned are cross-sectional, they do not answer the question of whether the obtained differences in personality-score levels are actually maturational-developmental or simply cohort effects. As has been so often stated, cross-sectional studies confound maturation with generational differences. "Generational differences," or "cohort effects" as they are also called, refer to the effects of the different socialization of successive birth cohorts. Resolution of this question of developmental versus cohort effects is of practical as well as theoretical importance. In a previous publication (Costa & McCrae, 1978b), we pointed out that cohort norms might be more appropriate and meaningful than age norms in the interpretation of certain test results.

To ascertain directly whether any consistent age differences observed were maturational rather than cohort or generational differences, we readministered, in 1975, the 16PF to a subsample of 139 men from the original 969 in the NAS. By measuring the same subjects at a later time (i.e., 10 years later), we hoped to generate data useful for resolving this question of maturational change in personality-trait levels (Costa & McCrae, 1978b).

Table 7.2 presents the results of repeated-measures analysis of variance (ANOVA) on all 16 scales. We used three age groups—25-40, 41-46, and over 47—as one of the two classifying variables, and administration as the second. Cross-sectional age differences, or cohort/aging differences, were found for Scales *G* (Conscientious), *I* (Tender-minded), and Q_1 (Liberal Thinking), all showing an increase across age groups. Two different scales, *B* (Bright) and Q_2 (Group-dependent), showed longitudinal changes (but not cohort differences), with increases over time. There were no significant time-by-cohort interactions. Because the conventional longitu-

TABLE 7.2. Group Means Averaged across Two Administrations

Scale		Age at first administration				
		25 to 40 ($n = 46$)	41 to 46 ($n = 51$)	47+ ($n = 42$)	*F* for age	*F* for time
A	(Outgoing)	−5.94	−5.28	−5.78	1.48	2.64
B	(Bright)	6.07	5.76	5.74	12.80	13.78***
C	(Stable)	7.04	5.91	6.50	1.93	.59
E	(Assertive)	−5.15	−5.25	−6.26	1.37	.36
F	(Happy-go-lucky)	−2.32	−1.54	−3.00	1.57	2.73
G	(Conscientious)	.47	1.06	1.92	5.49**	2.59
H	(Adventurous)	−1.17	−.52	−.49	.54	.21
I^a	(Tender-minded)	7.29	7.99	8.45	5.18**	1.27
L	(Suspicious)	−2.85	−2.69	−2.55	.30	2.94
M^a	(Imaginative)	11.98	12.03	12.53	1.92	2.48
N	(Shrewd)	1.97	2.42	2.63	2.96	.06
O	(Guilt-prone)	1.19	1.76	1.31	1.28	.05
Q_1^a	(Liberal Thinking)	9.37	9.81	10.08	3.35*	.96
Q_2	(Independent)	−3.40	−3.45	−2.92	.88	19.33***
Q_3	(Controlled)	1.08	59	1.07	1.83	.04
Q_4	(Tense)	−4.69	−3.91	−5.01	1.38	1.26

Note. Scales based on items common to 1962 and 1967 editions, Form A; scale norms do not apply.
[a] Based on full-scale retest in 1977, with $n = 134$, 101, and 169, respectively, for three age groups.
*$p < .05$.
**$p < .01$.
***$p < .001$.

dinal design we employed tested the same subjects at a later time and not a different subsample of the same age cohort, it is not possible to separate further the maturational changes from time-of-measurement or practice effects. Only by use of the sequential designs discussed later can this be approximated.

A simple maturational effect would be seen in both cross-sectional differences and longitudinal changes. That no scale showed this pattern of results leads to the suggestion that some other source of variation was responsible for the observed effects. The cross-sectional results can be attributed to cohort differences, but an explanation of the longitudinal findings is more difficult. It is possible that the changes were due to repeated exposure to the test or to cultural changes in the intervening 10 years. The change in Scale Q_2 is particularly difficult to understand, and since no comparable changes were found on any of the other scales in the Extraversion domain, perhaps the most parsimonious explanation is sampling error.

Yet another rival hypothesis can be mentioned as a likely explanation for the change in Scale *B* (Intelligence), which is, in fact, an ability scale rather than a personality scale. The 16PF was first administered in small groups, and implicit time pressure may have prevented some people from performing at their best. On the second administration, at home, individuals may have taken more time (or consulted the dictionary) and thus improved their scores. Thus, in addition to time-of-measurement and practice effects, we must be aware of changes in the assessment conditions or situations.

In the conventional longitudinal analysis of the GZTS scales by Douglas and Arenberg (1978), two aspects of the approach to measuring change are worthy of note. First, the authors presented an accounting of subject attrition, assessing the effects of subject loss by dividing their original sample ($n = 605$) into repeats (336 men who appeared in both the original sample and the longitudinal sample) and nonrepeats (269 men who appeared only in the original sample). Using a 2×3 (age groups 17–39, 40–59, 60–98) unweighted means ANOVA, significant main effects (differences) were found for the repeats versus the nonrepeats on four scales. In all age groups, men who were retested on the GZTS (repeats) were lower in Ascendance and higher in Objectivity, Friendliness, and Emotional Stability than nonrepeats. Although statistically significant, the differences were small in magnitude. Nevertheless, repeat subjects tended to be less neurotic (Emotional Stability, Objectivity, Friendliness) and less assertive.

The second notable aspect is that an attempt was made to determine whether the *magnitude of change* was related to age by computing part correlations between age and residual of the second measure adjusted for the first. Analyses of such longitudinal change scores showed overall decline in General Activity, Friendliness, Thoughtfulness, Personal Relations, and Masculinity. The magnitude of change was related to age for the General Activity scale, suggesting an accelerated change in this variable, but not for the other four scales that showed significant changes. Table 7.3 shows the longitudinal changes for the five scales that showed significant changes.

Thus, as the men in the Baltimore study aged approximately 7 years (ranging from 5.6 to 9.9 years on retest), their General Activity, or pace of activity, declined, as did their scores on Scales *F* (Friendliness), *T* (Thoughtfulness), *PR* (Personal Relations), and *M* (Masculinity). This last maturational change might be expected by theorists such as David Gutmann (1974) who hypothesize a sex-linked shift or crossover in sex roles with advancing age.

Sequential Evidence for Stability

As discussed earlier, conventional longitudinal analyses, whether of test–retest or repeated-measures variety, are susceptible to rival interpretations other than maturational effects, including time-of-measurement and practice effects. More adequate data-gathering and data-analytic designs involve the use of cross-sequential and time-sequential analyses, which provide a more comprehensive identification of change phenomena (Baltes, 1968; Baltes, Reese, & Nesselroade, 1977; Schaie, 1965, 1977).

Cross- and time-sequential analyses were performed on the GZTS scores for the Baltimore study sample by Douglas and Arenberg (1978). In cross-sequential analyses, independent samples of individuals born in the same historical period are compared at different times of testing. Since

TABLE 7.3. Longitudinal Changes from First to Second Administration on Five GZTS Scales

	Age decades at first administration					
GZTS scales	20s	30s	40s	50s	60s	70s
General Activity	1.67	−.10	−.01	−1.08	−2.04	−1.28
Friendliness	−.25	.00	−.88	−.44	−.07	−.92
Thoughtfulness	−2.33	−.78	−.54	−.07	−.48	−.65
Personal Relations	−.33	−1.13	−.38	−.78	−.74	−2.38
Masculinity	−.33	.14	−.61	−.61	.00	−.96

recruitment into the Baltimore study was continual, the Douglas and Arenberg study (as all our other sequential analyses) contrasted two successive intervals of testing (January 1958 through June 1968 with July 1968 to December 1974) rather than two distinct time points and thus only approximated a true cross-sequential design. Birth cohorts were defined in 8-year intervals from 1892–1899 to 1940–1947 in order to approximate the 7-year period between testing. Practice effects are eliminated in the cross-sequential design, as are main effects of cohort. But aging is confounded with time-of-measurement effects, and sampling differences may also be present.

In time-sequential analyses, independent samples of individuals of the same age are compared at different times of measurement. Maturation is ruled out, but any obtained differences (effects) may be due to either cohort or time of measurement. The second time-of-measurement sample in the cross-sequential analysis consisted of 238 men from Sample B (the later GZTS sample, described earlier) born during the same periods (i.e., 1892–1899, 1900–1907, 1908–1915, . . . , 1932–1939), but varying in age from 30 to 81 years. The second time-of-measurement sample in the time-sequential analyses again employed a subset of Sample B ($n = 240$), including men born between 1900 and 1947.

In the cross-sequential design, different-aged individuals from the same cohorts or historical periods are compared at different times of testing. The birth-cohort effect confounds cohort and aging and can be referred to as "cohort/aging." The time effect confounds secular changes between the times of measurement and aging and can be referred to as "time/aging." In the time-sequential design (independent samples of same-aged individuals, compared at different times of measurement), we can identify the effects as aging/cohort and time/cohort.

By comparing the consistency of results for different effects on various personality scales, one can make interpretations concerning the effects of maturation, generation, and cultural changes, although these interpretations are never unequivocal (see Adam, 1978). Table 7.4 summarizes the results of the sequential analyses along with the cross-sectional and longitudinal (repeated-measures) analyses of the GZTS scales. The only consistent maturational effect from the four types of analyses were found for Masculinity. Older age groups or cohorts in both samples showed declines in Masculinity over the seven decade groups from the 20s to the 80s. Within-subject changes over a 7-year time interval similarly showed declines for Masculinity. The absence of a time/cohort effect in the time-sequential analyses for Masculinity indicates that neither cohort differences nor cultural change from the first interval of measurement to the

TABLE 7.4. Summary of GZTS Results for Four Different Analyses

Scale	Cross-sectional cohort/aging[a]	Longitudinal time/practice/aging	Cross-sequential time/aging	Time-sequential time/cohort
General Activity	Declined (A,B)	Declined	NS	—
Restraint	Increased (A,B)	—	—	NS
Ascendance	Declined (A,B)	—	—	—
Sociability	Declined (A)	—	—	NS
Emotional Stability	No difference	—	Declined	Declined
Objectivity	No difference	—	—	—
Friendliness	Increased (A,B)	Declined	NS	Declined
Thoughtfulness	Increased (A)	Declined	—	Declined
Personal Relations	Increased (B)	Declined	Declined	Declined
Masculinity	Declined (A,B)	Declined	Declined	—

[a] A = effect significant in Sample A; B = effect significant in Sample B.

second accounts for the observed decreases in Masculinity. A small, but significant, decline in Masculinity found for the time/aging effect in the cross-sequential analysis confirms the interpretation that the observed cohort differences and the longitudinal declines are probably maturational in nature. Practice effects are unlikely to account for the longitudinal decline; in both the cross-sectional and sequential results, only first administration results are used.

General Activity score changes were interpreted as reflecting the operation of maturation since both cross-sectional and repeated-measures longitudinal results showed declines, and there were no significant time/cohort effects on the time-sequential analyses. The time/aging effect in the cross-sequential analysis showed a decrease, but it did not reach statistical significance. Thus the General Activity changes can be interpreted as maturational, but seem to occur predominately later in life.

The major import of these sophisticated, quasi-experimental statistical designs and analyses should be clearly recognized. None of the eight other personality traits (GZTS Scales *R, A, S, E, O, F, T,* and *P*) showed maturational changes. The *T, P,* and *F* scores reflected cultural changes and not maturational ones. The *R* and *A* scores reflected simple generational differences, with later born cohorts less restrained and more assertive than earlier born and therefore older cohorts.

We have devoted considerable time and space to these cross-sectional, longitudinal, and sequential analyses, so it is quite important that we not let the crucial facts slip from our grasp. For only two of the scales was the evidence consistent with maturational change in mean level of personality traits. Although many researchers are aware of the Douglas and Arenberg findings, it would appear that few clearly recognize the evidence for trait (mean-level) stability provided by their analyses and results. Even those scales that showed maturational effects changed very little. The magnitude of the maturational declines in General Activity and Masculinity amounts to about one-eighth of a standard deviation over a 7-year period, a change that is not at all of practical significance. In a somewhat lighter mood, we calculated the length of time it would take the average man in the Baltimore study sample to become "feminized," that is, to reach the mean Masculinity score of college females. At the rate of decline of .41 items every 6.6 years, it would take the older man (average age of 75) 136 years to become "feminized" under the admittedly preposterous assumption that one could live to 211 years!

How do the present results and conclusions compare with the data and judgments of other researchers in the field? After all, the studies reviewed deal with men only and with only two measurement points from 7 years (Baltimore study) to 10 years (NAS). Although the longitudinal and sequential-type analyses are quite scarce in comparison to the more numerous, if less informative, cross-sectional studies, there are two sequential studies—by Schaie and Parham (1976) and by Siegler and her associates at Duke—that bear examination. Since Siegler, George, and Okun (1979) used Form C of Cattell's 16PF in a sample of both women and men and with more than two measurement points, the results of their cross-sequential analyses of adult personality are of particular relevance. From a sample of 502 white, middle-class, adult subscribers to a health insurance plan in the Durham, North Carolina, area, 331 adults were assessed four times over an 8-year period (1968–1976). The authors formed 12 2-year cohorts born between 1899 and 1922, with the youngest cohort 46–47 years old, and the oldest 68–69 in 1969. Using a repeated-measures ANOVA design with 12 age cohorts, two sexes, and four times of measurement, the authors found that none of the 16 personality scales showed changes over time consistent with a maturational explanation. The only scale that showed a main effect for cohort and for time was the intelligence scale, Factor *B*, which seemed to be influenced by generational differences (later born cohorts score higher) and practice effects (later administrations yield higher scores) rather than aging, since the two effects are in opposite directions. There were main effects for sex on 5 of the 16 factors, which were in the sex-stereotyped direction and stable over time, and also two significant interactions. Our earlier longitudinal analysis of Form A (Costa & McCrae, 1978b) and the Siegler *et al.* (1979)

analysis of Form C of the 16PF are in rather substantial agreement, then, that there is little evidence for age-related change in personality in the adult and later years.

Schaie and Parham (1976) extracted 19 factors from a 75-item Social Responsibility scale and applied sequential methods to an analysis of change in these factors. Seventeen of the 19 factors showed stability, with only small changes in the other two factors. They concluded that "within the domain of factors identified in our study, we can with confidence support the stability model" (p. 152).

Individual Consistency over Time

Retest Data from Cattell's 16 Personality Factors Questionnaire

Retest correlations, or stability coefficients, assess the magnitude of personality consistency or change in the relative ordering of individuals, regardless of absolute level. These are among the most important analyses for longitudinal studies, for although different samples at different times, or different cohorts at one time, can be used to estimate age changes in trait levels, only repeated testing of the same individuals can speak to the degree of stability of individual differences.

Our first line of evidence for stability of personality came from an examination of longitudinal data from the Cattell 16PF, administered in 1966 and 1975, and from a short form of the EPI (the EPI-Q) devised by Floderus (1974), administered in 1976 (Costa & McCrae, 1977). Earlier work with cluster analyses of the 16PF had identified three clusters, two of which showed the same scale composition across all age groups. Retest correlations between the 16PF Neuroticism cluster scores over a 9-year interval in three age groups—25–34, 35–54, and 55–82—were .58, .67, and .69, respectively. The Extraversion–Introversion cluster scores showed even greater stability: .75, .70, and .84, respectively, for the young, middle-aged, and old groups. Even when we employed an alternate personality instrument—the 18-item EPI-Q—to measure Neuroticism and Extraversion in 1976, significant stability coefficients were obtained over a 10-year interval. Correlations of 16PF cluster scores in 1966 with EPI-Q measures in 1976 in the three different age groups were .41, .49, and .54 for Neuroticism, and .46, .54, and .53 for Extraversion measures. These "alternate-form," or equivalence-and-stability, coefficients are quite remarkable when one considers that the equivalence coefficients (which are a measure of the interchangeability of the tests measured contemporaneously) average only .56 for Neuroticism and .55 for Extraversion.

Enduring Dispositions in the Guilford-Zimmerman Temperament Survey

Recently we reported additional evidence directly related to the longitudinal stability of personality traits, with particular attention focused upon two specific hypotheses (Costa, McCrae, & Arenberg, 1980). The first hypothesis was that certain socially desirable traits, such as sociability, assertiveness, and others that define the broad domain of Extraversion, are particularly stable and that change is more characteristic of undesirable traits that might be interpreted as elements of the Neuroticism domain. The second hypothesis asserts that stability coefficients will increase with age. This hypothesis was based on evidence that stability coefficients increase for children (Nesselroade & Baltes, 1974) and on suggestions that personality is increasingly consolidated in old age (Neugarten, 1964).

We tested these hypotheses with the GZTS scale scores of the Baltimore study participants described earlier. The GZTS was administered to subjects approximately every 6 years. Because of complications in scheduling, a few subjects took the test 2 years in succession or failed to take the second retest. To maintain uniformity of time interval and number of administrations, longitudinal analyses were limited to subjects who took their second GZTS 5.0 to 7.9 years after their first ($M = 6.6$ years, $n = 460$) and to those who took their third GZTS 11.0 to 15.4 years after the first ($M = 12.9$, $n = 222$). Three age groups were formed: young (17–44 years, mean age 36.7, $n = 145$); middle (45–59 years, mean age 51.5, $n = 183$); and old (60–85 years, mean age 67.9, $n = 132$).

Table 7.5 gives the 6-year and the 12-year stability coefficients for the total samples, and the 12-year stability coefficients within each of the three age groups.

Under the hypothesis that stability should be greater in older age groups, one-tailed tests of the significance of differences between correlations were computed for each pair of age groups on each scale. Of the 30 comparisons at each interval, six were significant for the 6-year interval, three for the 12-year interval. Of these nine significant differences, four were in the predicted direction, five in the opposite direction. Not one of the specific findings for scales at the 6-year interval was replicated at the 12-year interval.

Finally, Table 7.6 gives the estimated reliability and stability of the "true" scores (Heise, 1969) in a subsample of 114 subjects of all ages who had complete data for all scales at all three times. Also in this table are simple retest coefficients for Times 1 to 2, 2 to 3, and 1 to 3. The similarity of correlations in the first and second 6-year periods is further evidence of the continuing stability of traits. It is also noteworthy that 6-year retest coefficients in this most select group are quite comparable to those

TABLE 7.5. Six- and 12-Year Retest Coefficients for GZTS Scales for Total Samples and 12-Year Retest Coefficients for Three Age Groups

Scale	6-year retest	12-year retest	12-year retest		
	Total sample (17 to 85 years)	Total sample (20 to 76 years)	Young group (20 to 44 years)	Middle group (45 to 59 years)	Old group (60 to 76 years)
General Activity	.83 (410)	.77 (192)	.77 (60)	.82 (93)	.78 (39)
Restraint	.71 (418)	.72 (193)	.61 (62)	.74 (94)	.76 (37)
Ascendance	.82 (401)	.83 (194)	.85 (62)	.85 (95)	.77 (37)
Sociability	.81 (393)	.74 (182)	.64 (62)[a]	.81 (88)	.66 (32)
Emotional Stability	.74 (427)	.70 (203)	.63 (68)	.76 (96)	.71 (39)
Objectivity	.71 (405)	.69 (191)	.66 (64)	.76 (87)	.59 (40)
Friendliness	.77 (418)	.74 (193)	.74 (64)[b]	.68 (88)[c]	.87 (41)
Thoughtfulness	.72 (418)	.73 (199)	.78 (64)	.71 (94)	.71 (41)
Personal Relations	.73 (385)	.68 (188)	.70 (62)	.64 (89)	.73 (37)
Masculinity	.75 (417)	.72 (200)	.73 (66)	.71 (94)	.70 (40)
Mean stability	.77	.73	.72	.75	.73

Note. n's are given in parentheses; numbers in parentheses in column headings are age range at Time 1. All correlations are significant at $p < .001$.

[a] Difference between young and middle groups is significant at $p < .05$.

[b] Difference between young and old groups is significant at $p < .05$.

[c] Difference between middle and old groups is significant at $p < .01$.

TABLE 7.6. Observed Retest Coefficients for Three Intervals and Estimated Reliability and Stability Coefficients for "True" Scores

Scale	Observed retest correlations			Estimated reliability	Estimated 12-year stability
	r_{12}	r_{23}	r_{13}		
General Activity	.83	.84	.80	.88	.92
Restraint	.75	.75	.71	.80	.89
Ascendance	.81	.85	.85	.82	1.00[a]
Sociability	.84	.82	.75	.91	.82
Emotional Stability	.77	.83	.71	.89	.80
Objectivity	.77	.82	.74	.86	.86
Friendliness	.81	.78	.77	.83	.93
Thoughtfulness	.73	.76	.71	.78	.91
Personal Relations	.70	.73	.68	.75	.91
Masculinity	.74	.77	.73	.79	.92

Note. Coefficients were calculated with formulas of Heise (1969) for a subsample of 114 men with complete data at three times.

[a] Observed 12-year retest coefficient is greater than estimated reliability.

presented in Table 7.1 for a more inclusive group of subjects who may have dropped out of the study after the second administration of the GZTS.

The retest coefficients presented here are among the highest in the literature for so long a period of time, comparable to the 2-week retest coefficients of many scales. When statistical corrections for unreliability are applied to obtain estimates of the "true" stability of the dispositions, even higher values are seen. These estimates, using Heise's (1969) formulas, are given in Table 7.6. Because they are only estimates, themselves subject to sampling error, anomalies such as the more-than-perfect stability of the Ascendance scale sometimes occur, and these coefficients must be regarded with due caution. Simple, uncorrected Pearson correlations are straightforward and familiar, and the correlations seen here reach impressively large magnitudes despite errors of measurement.

The question of increasing stability with age in adulthood seems to be clearly answered in the negative; 6- and 12-year stability coefficients are quite similar for three age cohorts whose members have an initial age range of nearly 70 years. Statistically significant differences in the magnitude of stability coefficients are scattered and inconsistent and do not support the hypothesis. Increased stability with age may be found among children and adolescents, but by young adulthood, stability in these dimensions of temperament is so high—near the limits of reliability of the instrument—that a ceiling effect diminishes the likelihood of any further increase in stability.

The question of differential stability for different traits is more difficult to answer since change and error of measurement are confounded in retest coefficients. It can be seen that, for the total sample, the three traits that constitute the extraverted factor of Social Activity—General Activity, Sociability, and Ascendance—have mean coefficients of .82 and .78 for 6- and 12-year intervals, respectively, whereas the neurotic traits of low Emotional Stability, Objectivity, Friendliness, and Personal Relations show corresponding coefficients of .74 and .70. The latter are certainly lower, accounting for only about three-fourths as much variance. The result is comparable to findings with the 16PF mentioned earlier, in which 9-year coefficients ranged from .70 to .84 for Extraversion and from .58 and .69 for Anxiety or Neuroticism. It may be the case that extraordinary stresses produce temporary neurotic tendencies in some people or that the distress attendant on neurotic traits leads some individuals to change, either by themselves or with the aid of friends or professional therapists.

On the other hand, it may also be the case that neuroticism is more difficult to measure reliably. The estimated stabilities of "true" scores, corrected for unreliability, are given in Table 7.6. These show an average value of .91 for the three Extraversion scales and .88 for the four Neuroticism scales, suggesting that there is little basis for inferring differential stability between these two domains of traits.

The data on which this study was based are taken from a single self-report instrument applied to a select male sample, and methodological artifacts, including sample selection and attrition, social desirability, and response sets, may have inflated the correlations. But a similar pattern of results in other samples using other objective measures (Costa & McCrae, 1977; Leon, Gillum, Gillum, & Gouze, 1979) as well as ratings (Block, 1971) argues that the results presented here cannot be dismissed wholly as method variance. Indeed, the theme of stability in personality is being heard with increasing frequency from a variety of sources. The conclusion seems to hold for women as well as men (Siegler *et al.*, 1979), for Germans as well as Americans (Grombach, 1976), and for adolescents as well older adults (Bachman, O'Malley, & Johnston, 1978).

Eliminating Response Bias as an Explanation for Stability

Response Styles and Age

Test Artifacts in Self-Report Measures

By far the greatest number of studies conducted on personality and aging have employed self-report instruments as the primary or sole source of data. This strategy has both advantages and disadvantages. On the positive side, objective self-report inventories have a number of properties that make them preferable to observer ratings or projective techniques. Self-report scales (except purely empirical ones) generally rest on the following rationale: If the individual is asked a number of related questions about his or her thoughts, feelings, or actions, he or she can usually be trusted to answer with some accuracy. Since the questions can cover typical behavior or quantify the frequency as well as the intensity of behavior or feelings, it is possible to sample the individual's experience widely. Because identical questions are asked of each respondent, and because each test is scored identically, individuals' scores are maximally comparable. These properties hold whether the scale is self-administered or read to the respondent by an interviewer. The former method has the advantage of allowing the fast and economical gathering of data, which, in addition to its practicality, has the scientific merit of encouraging large sample research and frequent replications. Finally, there are a number of phenomena of interest to psychologists, such as daydreaming and subjective well-being, for which the subject is the only reliable observer.

In contrast, projective tests and observer ratings are usually based on extremely limited samples of behavior and require extensive judgment on the part of the rater. The relationship between the observed behavior (say, inkblot responses) and the inferred characteristic (say, ego strength) often depends on a tortuous chain of reasoning, and clinicians may draw quite different inferences from the same signs. Unstructured interviews provide the rater with a different basis for rating each individual, and peer ratings traditionally have relied on the use of one or a few adjective pairs or ratings scales which may carry different meanings to different raters. For all these reasons, self-report instruments have generally shown a better record of internal consistency, retest reliability, and construct validity than have ratings or projective tests.[1]

But self-report inventories are also prone to certain problems. The transparency of the items makes it possible for individuals so motivated to present themselves favorably or unfavorably. The use of a standard format for answering questions (yes-no, a Likert scale, a rating bar) makes it possible for consistencies in the style of responding to distort the scores obtained from the instrument, leading to spuriously high consistency or correlations. Issues of social desirability, acquiescence, and extreme responding have been the source of interminable debate among personality psychologists, and no definitive resolution has been reached.

---

[1] It would be possible to incorporate the advantages of the self-report inventories with those of ratings by employing raters long familiar with the subject and by using a large number of standard questions to measure each construct. Research along these lines is currently being conducted.

The nihilistic interpretation of self-report responses as nothing but response sets (Berg, 1955) has been repeatedly answered (Block, 1965; Wiggins, 1966), and sophisticated test constructors have learned to balance scales in order to reduce acquiescence effects and to guard against social desirability in constructing and interpreting scales.

Longitudinal Analyses of Response Sets in the Guilford-Zimmerman Temperament Survey

There has, however, been relatively little research on age trends in response sets. Some writers (Schaie & Schaie, 1977) have argued persuasively that there have been enormous changes in the amounts and kinds of testing that individuals of different generations have been exposed to and that this may introduce unwanted sources of variance in tests. Increased cautiousness may alter the responses of older subjects (Botwinick, 1969), or standards of social desirability may change with age, bringing shifts in the influences of that set. Age changes or cohort differences in responding could account for observed differences in scale scores or could mask real changes that are occurring. To date, all of these are speculations, with little empirical foundation. Clearly, before any conclusions are drawn about aging and personality based on the use of objective personality measures, some information on these issues would be useful. Recent analyses of data from the Baltimore study contribute to a resolution.

Over the past 20 years, subjects in the Baltimore study have been given the GZTS every 6 years. Since recruitment into the study has been continuous, new samples of individuals, ranging in age from the 20s to the 90s, have been tested at a succession of times. By dividing ages, birth cohorts, and times of measurement into 6-year intervals, a variety of analytic designs may be applied to aid in the interpretation of changes or differences.

The men in the Baltimore study sample are community-dwelling volunteers, most with a background in science or the professions. Their commitment to a longitudinal study shows them to be more conscientious than average, and comparison of their scores on the GZTS with college norms show that they are better adjusted than that group. As volunteers, they have no particular incentive to falsify their scores, and thus the results of this study should not be hastily generalized to other testing situations (such as counseling or employment) where incentives to distort may be involved.

The GZTS was given to all subjects with standard instructions. Answer sheets provide three response options—"yes," "no," and "?"—but subjects were instructed to use the "?" option only if they were completely unable to select "yes" or "no." Following the suggestion of Guilford and Zimmerman (1949), scales containing more than three "?" responses were invalidated. This criterion has been used in all previously reported applications of the GZTS conducted by the Baltimore study. However, this exclusionary principle may distort results. In particular, if age produces caution in responding, a disproportionate number of older subjects may be excluded, perhaps especially the most cautious. Because of that possibility, a new approach was adopted in the analyses reported here. All the GZTS answer sheets were keypunched, so that responses to individual items could be analyzed. In the original scoring system, "?" responses were not scored and thus tended to lower the score of the individual. In the new system, responses were assigned a value of −1, 0, or +1, with the "?" represented as a neutral, rather than a negative, value.

The handbook for the GZTS (Guilford *et al.*, 1976) lists three scales that have been developed to estimate the influence of certain response sets: the Gross Falsification (GF) scale, the Subtle Falsification (SF) scale, and the Careless Deviancy (CD) scale. The first two are intended to screen individuals who may be attempting to present an unduly favorable impression; the third is a scale composed of relatively rare responses, and a high score is interpretable either as careless responding or deviancy in personality. In addition, it is possible to measure at least three other response sets on the GZTS. The number of blanks was summed across all scales to give an index, as was the number of "?" responses. Items in the GZTS are roughly balanced on most scales, so it is possible to sum up the number of "yes" responses and consider it an index, not of any substantive personality trait, but of the tendency to acquiesce to items indiscriminately.

For each of these variables, cross-sectional, longitudinal, cross-sequential, and time-sequential analyses were performed. In the repeated-measures analysis, 348 men ranging in initial age from 32 to 74 and retested after 4 to 8 years (mean interval = 6.6 years) were the subjects. They were classified into seven age groups, each spanning a 6-year interval. In the time-sequential analyses, 328 men who were tested in the period from 1958 to 1964 were compared with 278 men tested between 1965 and 1971. They were cross-classified by the same seven age categories used in the repeated-measures design. In the cross-sequential analysis, 345 men tested between 1958 and 1964 were compared with 285 men tested between 1965 and 1971. In this design, however, they were cross-classified according to their dates of birth, using seven birth cohorts of 6-year intervals from 1896 to 1932.

Repeated-measures analyses showed a significant ($p < .05$) effect on the repeated factor for number of question marks, which increased from 6.2 to 9.6; on acquiescence, which decreased from 132.1 to 128.9; on the

GF scale, which increased from 11.8 to 12.1; and on the SF scale, which decreased from 21.2 to 20.8. Age-group differences were seen for the number of question marks, which were highest in the 68–74-year-old group and lowest in the 50–55-year-old group. In addition, there were two interactions: Men aged 38–43 at first testing showed a decrease instead of an increase in question marks, and men aged 55–61 showed an increase instead of a decrease in acquiescence.

These results are somewhat puzzling; certainly they did not show a monotonic change in any of the response sets with age. Most of the changes were extremely small in magnitude, and if we require a significance level of $p < .01$, only two effects are significant: the increase in question marks and the decrease in acquiescence. Neither of these longitudinal changes was mirrored in cross-sectional differences, suggesting that the changes are due either to time-of-measurement effects (i.e., a cultural change during the testing period) or to "practice" (i.e., repeated exposure to the test).

Examination of the cross- and time-sequential analyses, conducted on samples of more than 600 men, is revealing. Analyses of number of blanks, number of question marks, acquiescence, GF, SF, and careless deviancy show *no* significant ($p < .05$) effects for aging/cohort, aging/time, cohort/time, or cohort/aging, nor were there any significant interactions. These data suggest that the marginal cross-sectional differences and interactions in the repeated-measures analyses are best regarded as unreplicable error and that the longitudinal changes in acquiescence and use of question marks are attributable to practice effects. That acquiescence decreased while use of question marks increased by three items is suggestive: Perhaps subjects felt pressured on the first administration to avoid question marks at all costs and agreed to a few items of which they were uncertain. Some years later, as experienced subjects no longer as reticent to assert themselves, they used the question marks when they felt they needed to.

In any case, these data imply that response sets are not ordered by age. The longitudinal researcher may want to consider the effects of repeated administration of the same instrument, but the particular effect seen here is small in its overall influence on scale scores and probably is unique to instruments like the GZTS, which provide a question-mark option, but fail to score it.

Response Sets in the Neuroticism–Extraversion–Openness Inventory

But what about questionnaires that use a different response format? Likert scales, for example, are subject to the influence of extreme responding—the tendency to "strongly agree" or "strongly disagree" to a variety of questions. And are aging women perhaps more susceptible to response sets than aging men? We have no longitudinal data on which to base answers to these questions, but cross-sectional data on a new personality instrument, the NEO Inventory, do provide some indirect evidence of the effect of age on response styles in the Likert format for men and women.

The Neuroticism-Extraversion-Openness (NEO) Inventory consists of 144 items measuring three broad domains of personality—Neuroticism, Extraversion, and Openness to Experience—and six more specific facets within each of these. Subjects are asked to respond to each item on a 5-point scale from "strongly disagree" to "strongly agree." They are instructed to use the middle, neutral category if they are undecided. The scales are balanced, so a count of the number of items to which a subject responds "agree" or "strongly agree" can be taken as a content-free measure of acquiescence. A "naysayer" index can be created by counting the "disagree" and "strongly disagree" responses. Neutral responses can also be counted. Finally, it is possible to count the number of extreme responses—"strongly disagree" and "strongly agree"—in order to index extreme responding.

Using the same 6-year age groups as were used previously, these scores were analyzed for a sample of 265 men and 191 women. There were no significant effects for age group or gender, nor were there any significant interactions. Once again, the evidence suggests that stability or change in mean level of self-report personality inventories is not likely to be an artifact of response sets. It also suggests that older men and women respond to questionnaires in much the same way as younger people do. These results are but one of many pieces of evidence that must be considered in evaluating the validity of self-report inventories when used with the elderly. The construct validity, factorial invariance, and predictive utility of trait measures discussed in other sections of this chapter also contribute to the conclusion that this approach to aging and personality has much to offer.

Retest Stability and Response Bias

We have shown that such extraneous sources of variance as acquiescence, extreme responding, and social desirability are not themselves age related. But it might be argued that the high level of stability of individual differences in personality dispositions is, in fact, better regarded as the stability of responding styles. If the influence of these variables were removed, would the stability coefficients be as high?

To test this, we calculated the retest coefficients of the GZTS, partialling out the variance due to response sets. Clearly, the response scales derived from the GZTS itself cannot be used for this since they are based on the same items as the temperament scales. If the individual endorses the same item on both occasions, we cannot tell whether that should be attributed to stability of trait or response style. An independent measure of response bias, however, could be used in this analysis, and the response-style indexes derived from the NEO Inventory just described were available. As a measure of social desirability or deliberate falsification, the Lie scale of the EPI was also available. Unfortunately, only 98 subjects had complete data on all these variables, so there is a substantial difference in sample from the original stability data.

Table 7.7 gives the retest coefficients for the ten GZTS scales over an interval of from 4 to 8 years (mean = 6.6 years), statistically controlling for acquiescence, naysaying, and extreme responding as measured by the response indexes of the NEO, and for falsification as measured by the EPI Lie scale. It should be noted that data on the response-set variables were collected in 1980, in some cases several years after the second administration of the GZTS. However, since the hypothesis under consideration is that response styles account for the observed stability of personality-scale scores, these styles must, by hypothesis, be themselves extremely stable. The point at which they are measured is therefore more or less immaterial.

An examination of this table shows that most of the stability coefficients

TABLE 7.7. Retest Correlations of GZTS Scales over a 6-Year Interval Partialing out the Effects of Acquiescence, Naysaying, Extreme Responding, and Falsification in a Sample of 98 Men

Scale	Corrected correlation	Partial correlation
General Activity	.85	.85
Restraint	.71	.70
Ascendance	.83	.83
Sociability	.83	.82
Emotional Stability	.73	.72
Objectivity	.66	.64
Friendliness	.74	.74
Thoughtfulness	.73	.71
Personal Relations	.73	.74
Masculinity	.73	.68

are unchanged after removing the effects of response set. Only Masculinity showed a noticeable change, and this can be attributed to a small correlation (−.31) between Masculinity and extreme responding at each time. The statistical procedure of partial correlation merely demonstrates a fact that is evident when the simple correlations are examined: None of the response sets is substantially related to any of the GZTS scales, and it is therefore not the case that stability in these response styles could account for stability in personality-scale scores. Incidentally, this is also additional evidence for the validity of the GZTS scales.

Personality Variables in Longitudinal Research

Eliminating the Effects of Mood States

One of the great advantages of many longitudinal studies is the accumulation of data from a variety of measures on the same set of subjects. Only from a study in which data collection is carried out over a period of years is it feasible to measure health, cognitive ability, personality, life events, well-being, perception, and so on, in a sufficiently large sample to allow multivariate analyses of all these variables. This wealth of data is one of the primary rationales for conducting longitudinal studies and explains why so many of them are interdisciplinary in scope.

But longitudinal studies also incorporate time intervals between the measures collected, and the imaginative use of data collected at different times can contribute enormously to an understanding of causal sequences and developmental progressions. Indeed, the study of stability or change in a single measured variable is only the simplest and most basic kind of longitudinal analysis. Many more interesting variations are possible.

One major use of longitudinal analysis is seen in the elimination of certain time-of-measurement effects. In personality and some kinds of cognitive testing, the transient state of the individual may influence the results of testing. Test conditions, fatigue, or a temporary dysphoric mood may contaminate the measurement of personality traits that are presumed to be enduring dispositions. Such influences not only would distort the measurement of the trait, but also might wholly account for observed relationships between variables measured at the same time and under the same conditions.

Consider the relationship between personality and subjective well-being. Since Wilson's (1967) review, it has been known that happiness, or subjective well-being, was correlated with personality. Contemporaneous studies using personality measures from the Buss and Plomin EASI-III Temperament Survey, EPI, and the 16PF, and well being measures from the Bradburn Affect Balance Scales, the Knutson Personal Security Inventory, the Beck Hopelessness Scale, and an index of life satisfaction showed clearly that Extraversion was associated with happiness, and especially with the Positive Affect scale, and that Neuroticism was negatively related to happiness, and especially the Negative Affect scale (Costa & McCrae, 1980a). Using this variety of measures and sample sizes of several hundred men, it was possible to conclude that the relationship was not an artifact of the particular measures used. But the causal interpretation, that personality influenced or contributed to the determination of happiness, could not be sustained on the basis of concurrent correlations alone. It could be argued that the mood of the individual at the time of the administration of the test might account for the observed correlation—that the responses to a personality test of a happy man resembled those of an extravert, whereas the responses of an unhappy man mimicked neuroticism.

However, 10-year longitudinal data speak directly to this question. If personality characteristics measured 10 years previously show significant relationships to well-being 10 years later, the causal attribution is much more easily made. In fact, we found that 16PF Extraversion scores were related to Bradburn's Affect Balance ($r = .14$, $p < .05$) and Positive Affect ($r = .23$, $p < .001$), but not to Negative Affect ($r = .03$, NS). The 16PF Neuroticism scores, also as predicted, were related to Affect Balance ($r = -.36$, $p < .001$) and to Negative Affect ($r = .39$, $p < .001$), but not to Positive Affect ($r = -.08$, NS). Analyses conducted on the same data set subsequent to the publication of these results show that, over a 10-year interval, 16PF Neuroticism also predicts Hopelessness ($r = .40$, $n = 376$, $p < .001$) and Personal Security ($r = -.38$, $n = 258$, $p < .001$). Corresponding correlations for 16PF Extraversion scores were $r = -.19$, $p < .001$, and $r = .33$, $p < .001$, respectively.

These findings were replicated using a different sample and different measures of both dependent and independent variables (Costa, McCrae, & Norris, 1981). In this study, personal adjustment to aging as measured by scales from the Chicago Attitude Inventory was predicted from Extraversion and Neuroticism factors of the GZTS. In addition to contemporaneous associations, GZTS factors were also shown to predict personal adjustment 6 and 12 years later.

Predicting across Periods of the Life Span

It is also possible that there are unwanted sources of variance that contaminate all measures taken in an entire period of time. A stressful life event, such as the loss of a spouse, may influence personality-inventory responses, mood measures, coping styles, and even cognitive performance for weeks or months. Even if researchers have gone to the trouble of measuring variables on different days or in different testing conditions, they cannot rule out the possibility of this kind of effect without employing a sufficiently long time lag.

An instance of this is provided by two analyses of the male midlife crisis (Cooper, 1977; Costa & McCrae, 1978b). In these two studies, a self-report measure of characteristics and concerns hypothesized to contitute the "midlife crisis" (the MLC scale) was found to have no relationship to age, but a marked correlation was observed between the MLC scale and the Eysenck Neuroticism scale administered at about the same time. One interpretation was that only men high in neuroticism suffered a midlife crisis, but an alternative was offered by developmental-stage theorists. Levinson had written that "because a man in this crisis is often somewhat irrational, others may regard him as 'upset' or 'sick.' In most cases, he is not" (Levinson *et al.*, 1978, p. 199). Such individuals are supposed to be in the midst of "tumultuous struggles," in which they call into question their entire lives and are "horrified by much that is revealed. They are full of recriminations against themselves and others" (p. 199). It is entirely reasonable to suppose that a normal individual going through such a transition would appear high on a scale of neuroticism. That these persons did not cluster anywhere near the predicted age for a midlife crisis leads, however, to some skepticism, and the availability of longitudinal personality data made possible a clear test of these rival hypotheses. If the crisis were a continuation of long-standing maladjustment, we would expect a correlation between earlier measures of neuroticism and MLC scale scores. If the contemporaneous correlation were a distortion, caused by the developmental crisis of the midlife transition, then there should be no relationship—indeed, a Jungian theorist might even hypothesize that those apparently best adjusted in early adulthood might have the most difficult readjustment at midlife and that a negative correlation should be expected.

Using 16PF data collected 10 years previously, we showed significant positive correlations between the MLC scale and each of the five scales loading on the General Anxiety or Neuroticism cluster. Individuals who will one day experience something that is phenomenologically similar to a midlife crisis are those with a long history of maladjustment. It is possible that at certain times their emotional stability is particularly low, so contemporaneous correlations may somewhat inflate the actual relationship, but these longitudinal data provide compelling evidence that the midlife crisis is more a matter of neuroticism than of development.

Cause and Effect in Psychosomatic Research

The analysis just suggested, in which a predictive relationship is compared to a concurrent one, is chiefly of interest as a means of eliminating the possible effects of transient states and is usually based on the assumption that the true variable is constant and that variations from one time to another are the result of errors in measurement. It is also possible to entertain the possibility that real changes may occur in the variable. Consider the case of coronary heart disease (CHD). The psychic shock of discovering that one has heart disease, and the alterations in activity and life-style that the individual may be asked by his or her physician to make, could plausibly alter enduring personality dispositions. For that reason, research reporting associations of personality with any phase of CHD are inherently ambiguous.

In some preliminary results, it was noted that Baltimore study subjects who complained of angina were lower than average on two of the GZTS scales: Emotional Stability and Masculinity. Was this an example of personality change resulting from medical illness, or were personality variables diagnostic of and perhaps causally involved in the development of the disease? To tease out these possibilities, it was necessary to separate subjects into different groups and to take into consideration the temporal course of their disease (Costa, Flegg, Lakata, & McCrae, 1980). Eighty-eight subjects were measured on the GZTS during their first or second visit to the Gerontology Research Center, and all were free from both anginal complaints and certain ECG signs on these visits. Over the next 20 years, four groups emerged: those who developed CHD as evidenced by both anginal complaints and ECG signs; those who showed ECG signs of CHD, but who never reported angina in a follow-up period of from 5 to 15 years; those who reported angina, but who showed no ECG signs of CHD in the same follow-up period; and an age-matched control group (age-matched to the total of the three groups) that showed neither ECG signs nor angina in a follow-up period of 10–20 years. (All subjects were classified according to their status at last examination; the varying follow-up intervals reflect the fact that subjects entered and left the study at different times.)

The first group is easily diagnosed as having CHD. The second and third groups are more ambiguous: Those with only angina seem to be overly sensitive; those with only ECG signs appear undersensitive. Do any of these distinctions show up in personality measures taken before the development of disease?

Our results showed that there were no differences between the first and fourth groups, that is, between those who definitely did and did not develop CHD. Thus none of the traits measured by the GZTS appears to have etiological significance. But there were preexisting differences between the two intermediate groups. Individuals who complained of anginal pains, but who gave no ECG evidence of disease, were lower on Masculinity and less emotionally stable than those who reported no angina despite ischemic ECG signs.

It is well known (Hurst, Logue, Schlant, & Wenger, 1978; Froelicher, 1977) that resting and stress ECG signs in themselves are far-from-perfect indicators of CHD, and the number of subjects in this study (88) is too small to be conclusive. Nevertheless, some interesting interpretations are suggested by the data. Individuals (or at least men) high in Masculinity and Emotional Stability may be more likely to minimize minor chest pains, and the discovery of disease in them may depend on routine medical procedures. Without routine examinations, these men are likely to remain undiagnosed and untreated (cf. Berglund, Ander, Lindstrom, & Tibblin, 1975). On the other hand, men who are low in Masculinity and Emotional Stability are very sensitive to chest pains and may report anginal symptoms even in the absence of organic pathology. These people may request unnecessary medical attention. Taken together, these analyses suggest that the personality variables measured by the GZTS do not affect the development of CHD, but they do seem to affect the presentation of symptoms.

If these personality differences were found at the same time as the symptoms, several alternate interpretations could be offered. We might argue that the experience of angina was sufficiently traumatic to lower the individual's emotional stability and that the restraints on behavior, imposed by the patient or his physician, had diminished his sense of masculinity. Or we might argue that self-selection played the crucial role: Those individuals who were aware of chest pains and who were also predisposed to worry about their health (i.e., the more neurotic) would be most likely to join and remain in a longitudinal study that promised periodic monitoring of their health. Better adjusted men with angina, some of whom could be expected to show no ECG signs and thus be classified in the third group, would be less concerned with their health and less likely to volunteer for the study. These kinds of arguments have been supported in studies of the relationship between hypertension and personality (Costa, McCrae, Andres, & Tobin, 1980).

But in this case, we can rule out those alternatives. The measurement of personality preceded the development of angina and ECG signs, so it could not have been influenced by them. Likewise, self-selection on this basis is impossible, since the individuals were presumably unaware that they would experience angina in the next few years. Retrospective accounts of personality difference could not be trusted in this context; only the archives of a longitudinal study permit the kinds of inferences offered here.

These rather simple examples of the kinds of inferences that can be ruled out with longitudinal data are useful in part because they illustrate the logic involved. Path-analytic techniques, however, offer far more sophisticated statistical models, which estimate not only the direction of causal influence, but the degree of influence, expressed as a regression weight (see Kenny, 1979). Causal inferences based on these statistical techniques are subject to a number of restrictions, many of them related to the assumptions necessary to constructing the model. In general, fewer assumptions and more reliable results can be obtained by having more measurement points and by measuring more of the variables that might plausibly have an influence on elements in the model. Clearly, longitudinal studies are ideal for this type of analysis.

Conclusions

The Utility of a Trait Model

A chapter on longitudinal research on personality is not the place for a defense of self-report measures, the trait approach, or a particular model of personality dimensions. But since we began with a disclaimer alerting the reader to the wide variety of traditions in personality that are not reviewed in this chapter, it may be useful, in closing, to review the benefits of the approach we have adopted.

Within the scope of this chapter, we have shown that traits in the domains of Neuroticism, Extraversion, and Openness to Experience are indeed the "enduring dispositions" posited by trait theorists. This is a major contribution to the field of personality that only longitudinal studies could make. We have addressed the issue of the possible influence of response sets on stability of personality and have found that it is negligible. This helps to restore confidence in self-report methods, which in practice have always dominated empirical personality research. By aggregating traits into covarying clusters or domains, we have been able to structure a review of the literature and to formulate generalizations that may apply to many trait measures not hitherto used in longitudinal research. Further research will judge the adequacy of these generalizations. Finally, in the application to such problems as subjective well-being, the midlife crisis, and certain kinds of symptom reports, we have demonstrated the value of our trait model in predicting outcomes or criteria of interest to students of adulthood and aging.

Considerations of convenience contributed to the choice of implementing objective personality inventories in longitudinal studies, but the choice has turned out to be fortunate. Yet it must be remembered that the value of one approach does not imply the uselessness of others. Large-scale longitudinal studies of nontrait variables, such as ego functions or coping processes used to deal with chronic and acute stress, should be conducted. Perhaps the most fruitful longitudinal research will study such processes across the life span in conjunction with enduring personality dispositions.

Research on Personality and the Life Course

The major drawback to the use of a stable individual-difference approach to life-span development is that current conceptions and methods of research on aging and personality are designed primarily for the study of change. The elegant models for attempting to separate true maturational changes from cohort differences and cultural changes are not very useful if there is no meaningful maturational change. The widespread attempt to chart the developmental course of personality in adulthood no longer seems profitable, and it may not be immediately clear what direction future research should take.

We have argued elsewhere (McCrae & Costa, 1982) that personality dimensions may be more usefully construed as causes than as effects. Age, and its attendent social, cognitive, and biological changes, should be considered in conjunction with personality as joint shapers of the life course. Personality can help explain the choices (educational, career, familial) that must be made at specific life transitions. The stability of personality dispositions may contribute to the individual's sense of identity and to the continuity and coherence of the life course. Finally, adaptation to life at all ages is likely to be powerfully influenced by personality.

Prospective life histories collected by longitudinal studies can help to answer such questions as the following: How do the lives of introverts differ from the lives of extraverts? Does openness to experience lead to a more fluid and unpredictable life course? How is neuroticism typically manifested at each stage of life? Answers to these questions may help in integrating the insights of life-span developmental psychology with those of personality in the study of lives.

References

Adam, J. Sequential strategies and the separation of age, cohort, and time-of-measurement contributions to developmental data. *Psychological Bulletin*, 1978, *85*, 1309-1316.

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.

Allport, G. W. Traits revisited. *American Psychologist*, 1966, *21*, 1-10.

Allport, G. W., & Allport, F. H. *A-S reaction study*. Boston: Houghton Mifflin, 1928.

Bachman, J. G., O'Malley, P. M., & Johnston, J. *Adolescence to adulthood: Change and stability in the lives of young men*. Ann Arbor, Mich.: Institute for Social Research, 1978.

Baltes, P. B. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, 1968, *11*, 145-171.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. *Life-span developmental psychology: Introduction to research methods*. Monterey, Calif.: Brooks/Cole, 1977.

Beck, A. T. *Depression: Causes and treatment*. Philadelphia: University of Pennsylvania Press, 1972.

Bendig, A. W. Age differences in the interscale factor structure of the Guilford-Zimmerman Temperament Survey. *Journal of Consulting Psychology*, 1960, *24*, 134-138.

Berg, I. A. Response bias and personality: The deviation hypothesis. *Journal of Psychology*, 1955, *40*, 61-71.

Berglund, G., Ander, S., Lindstrom, B., & Tibblin, G. Personality and reporting of symptoms in normo- and hypertensive 50-year-old males. *Journal of Psychosomatic Research*, 1975, *19*, 139-145.

Block, J. *The challenge of response sets*. New York: Appleton-Century-Crofts, 1965.

Block, J. *Lives through time*. Berkeley, Calif.: Bancroft Books, 1971.

Block, J. Advancing the psychology of personality: Paradigmatic shift or improving the quality of research. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Botwinick, J. Disinclination to venture response versus cautiousness in responding: Age differences. *Journal of Genetic Psychology*, 1969, *115*, 55-62.

Buss, A. H., & Plomin, R. *A temperament theory of personality development*. New York: Wiley, 1975.

Byrne, D. Repression-Sensitization as a dimension of personality. In B. A. Maher (Ed.), *Progress in experimental personality research* (Vol. 1). New York: Academic Press, 1964.

Cattell, R. B. *Personality and mood by questionnaire*. San Francisco: Jossey-Bass, 1973.

Cooper, M. W. *An empirical investigation of the male midlife period: A descriptive, cohort study*. Unpublished undergraduate honors thesis, University of Massachusetts at Boston, 1977.

Costa, P. T., Jr., & McCrae, R. R. Age differences in personality structure: A cluster analytic approach. *Journal of Gerontoloy*, 1976, *31*, 564-570.

Costa, P. T., Jr., & McCrae, R. R. Age differences in personality structure revisited: Studies in validity, stability, and change. *Aging and Human Development*, 1977, *8*, 261-275.

Costa, P. T., Jr., & McCrae, R. R. Letter to the editor. *Journal of Gerontology*, 1978, *33*, 4. (a)

Costa, P. T., Jr., & McCrae, R. R. Objective personality assessment. In M. Storandt, I. C. Siegler, & M. F. Elias (Eds.), *The clinical psychology of aging*. New York: Plenum, 1978. (b)

Costa, P. T., Jr., & McCrae, R. R. The influence of extraversion and neuroticism on subjective well-being. Happy and unhappy people. *Journal of Personality and Social Psychology*, 1980, *38*, 668-678. (a)

Costa, P. T., Jr., & McCrae, R. R. Somatic complaints in males as a function of age and neuroticism: A longitudinal analysis. *Journal of Behavioral Medicine*, 1980, *3*, 245-257. (b)

Costa, P. T., Jr., & McCrae, R. R. Still stable after all these years: Personality as a key to some issues in adulthood and old age. In P. B. Baltes & O. G. Brim (Eds.), *Life span development and behavior* (Vol. 3). New York: Academic Press, 1980. (c)

Costa, P. T., Jr., McCrae, R. R., Andres, R., & Tobin, J. D. Hypertension, somatic complaints and personality. In M. F. Elias & D. Streeten (Eds.), *Hypertension and cognitive processes.* Mt. Desert, Me.: Beech Hill Publishing, 1980.

Costa, P. T., Jr., McCrae, R. R., & Arenberg, D. Enduring dispositions in adult males. *Journal of Personality and Social Psychology*, 1980, *38*, 793–800.

Costa, P. T., Jr., McCrae, R. R., & Norris, A. Personal adjustment to aging: Longitudinal prediction from Neuroticism and Extraversion. *Journal of Gerontology*, 1981, *36*, 78–85.

Costa, P. T., Jr., Flegg, J., Lakata, E. & McCrae, R. R. *Longitudinal personality predictors of anginal complaints in males.* Paper presented at the 33rd Annual Meeting of the Gerontological Society, San Diego, Calif., November 1980.

Cunningham, W. R. Principles for identifying structural differences: Some methodological issues related to comparative factor analyses. *Journal of Gerontology*, 1978, *33*, 82–86.

Douglas, K., & Arenberg, D. Age changes, cohort differences, and cultural change on the Guilford-Zimmerman Temperament Survey. *Journal of Gerontology*, 1978, *33*, 737–747.

Eysenck, H. J. *The structure of human personality.* London: Methuen, 1960.

Eysenck, H. J., & Eysenck, S. B. G. *Manual of the Eysenck Personality Inventory.* San Diego, Calif.: Educational and Industrial Testing Service, 1975.

Floderus, B. Psycho-social factors in relation to coronary heart disease and associated risk factors. *Nordisk Hygienisk Tidskrift Supplementum 6*, Stockholm, Sweden, 1974.

Fozard, J. L., & Thomas, J. C. Psychology of aging: Basic findings and their psychiatric applications. In J. G. Howells (Ed.), *Modern perspectives in the psychiatry of old age.* New York: Brunner-Mazel, 1975.

Froelicher, V. F. Use of the exercise electrocardiogram to identify latent coronary atherosclerotic heart disease. In E. A. Amsterdam, J. H. Wilmore, & A. N. DeMaria (Eds.), *Exercise in cardiovascular health and disease.* New York: Yorke Medical Books, 1977.

Gorsuch, R. L. *Factor analysis.* Philadelphia: W. B. Saunders, 1974.

Gould, R. L. *Transformations.* New York: Simon & Schuster, 1978.

Grombach, H. Consistency and change of personality variables in late life. In H. Thomae (Ed.), *Patterns of aging: Findings from the Bonn Longitudinal Study of Aging.* Basel, Switzerland: S. Karger, 1976.

Guilford, J. P., & Zimmerman, W. S. *The Guilford-Zimmerman Temperament Survey: Manual of instructions and interpretations.* Beverly Hills, Calif.: Sheridan Supply Company, 1949.

Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. *The Guilford-Zimmerman Temperament Survey handbook: Twenty-five years of research and application.* San Diego, Calif.: EdITS Publishers, 1976.

Gutmann, D. Alternatives to disengagement: The old men of the Highland Druze. In R. Levine (Ed.), *Culture and personality: Contemporary readings.* Chicago: Aldine, 1974.

Gynther, M. D. Aging and personality. In J. N. Butcher (Ed.), *New directions in MMPI research.* Minneapolis: University of Minnesota Press, 1980.

Heise, D. R. Separating reliability and stability in test–retest correlation. *American Sociological Review*, 1969, *34*, 93–101.

Hurst, J. W., Logue, R. B., Schlant, R. C., & Wenger, N. K. (Eds.). *The heart.* New York: McGraw-Hill, 1978.

Jung, C. G. *Psychological types.* New York: Harcourt, Brace & World, 1933.

Kenny, D. A. *Correlation and causality.* New York: Wiley, 1979.

Lawton, M. P., Whelihan, W. M., & Belsky, J. K. Personality tests and their uses with older adults. In J. Birren (Ed.), *Handbook of mental health and aging.* New York: Prentice-Hall, 1980.

Leon, G. R., Gillum, B., Gillum, R., & Gouze, M. Personality stability and change over a 30 year period—Middle age to old age. *Journal of Consulting and Clinical Psychology*, 1979, *23*, 245–259.

Levinson, D. J., Darrow, C. N., Klein, E. B., Levinson, M. H., & McKee, B. *The seasons of a man's life*. New York: Alfred A. Knopf, 1978.

McCrae, R. R. & Costa, P. T., Jr. Aging, the life course, and models of personality. In T. Field (Ed.), *Review of human development*. New York: Wiley, 1982.

McCrae, R. R., Costa, P. T., Jr., & Arenberg, D. Constancy of adult personality structure in males: Longitudinal, cross-sectional and times-of-measurement analyses. *Journal of Gerontology*, 1980, *35*, 877–883.

Nesselroade, J. R., & Baltes, P. B. Adolescent personality development and historical change: 1970–1972. *Monographs of the Society for Research in Child Development*, 1974, *39* (Serial No. 154).

Neugarten, B. L. *Personality in middle and late life*. New York: Atherton, 1964.

Rokeach, M. *The open and closed mind*. New York: Basic Books, 1960.

Schaie, K. W. The effect of age on a scale of social responsibility. *Journal of Social Psychology*, 1959, *50*, 221–224.

Schaie, K. W. A general model for the study of developmental problems. *Psychological Bulletin*, 1965, *64*, 92–107.

Schaie, K. W. Quasi-experimental research designs in the psychology of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold, 1977.

Schaie, K. W., & Parham, I. A. Stability of adult personality: Fact or fable. *Journal of Personality and Social Psychology*, 1976, *34*, 146–158.

Schaie, K. W., & Schaie, J. Clinical assessment and aging. In J. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold, 1977.

Sealy, A. P., & Cattell, R. B. *Standard trends in personality development in men and women of 16 to 70 years, determined by 16PF measurements*. Paper presented at the British Social Psychology Conference, London, April 1965.

Siegler, I. C., George, L. K., & Okun, M. A. Cross-sequential analysis of adult personality. *Developmental Psychology*, 1979, *15*, 350–351.

Spielberger, C. D. Anxiety as an emotional state. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 1). New York: Academic Press, 1972.

Tellegen, A., & Atkinson, G. Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, 1974, *83*, 268–277.

Thomae, H. Personality and adjustment to aging. In J. E. Birren & R. B. Sloane (Eds.), *Handbook of mental health and aging*. Englewood Cliffs, N.J.: Prentice-Hall, 1980.

Wagner, E. E. Differences between old and young executives on objective psychological test variables. *Journal of Gerontology*, 1960, *15*, 296–299.

Wiggins, J. S. Substantive dimensions of self-report in the MMPI item pool. *Psychological Monographs*, 1966, *80* (22, Whole No. 630).

Wilson, W. Correlates of avowed happiness. *Psychological Bulletin*, 1967, *67*, 294–306.

Zemore, R., & Eames, N. Psychic and somatic symptoms of depression among young adults, institutionalized aged and noninstitutionalized aged. *Journal of Gerontology*, 1979, *34*, 716–722.

Zuckerman, M. *Sensation seeking: Beyond the optimal level of arousal*. Hillsdale, N.J.: Erlbaum, 1979.

Journal of Gerontology
Vol. 36, No. 1, 78-85

Personal Adjustment to Aging: Longitudinal Prediction from Neuroticism and Extraversion

Paul T. Costa, Jr., PhD, Robert R. McCrae, PhD,
and Arthur H. Norris, MS[1]

Personal adjustment to aging as measured by scales from the Chicago Attitude Inventory (CAI) was examined longitudinally in a community-dwelling sample of 557 men aged 17 to 97. Concurrent and predictive relations between this age-appropriate measure of well-being and personality were examined by correlating the CAI variables with three factors from the Guilford-Zimmerman Temperament Survey identified as Neuroticism, Extraversion, and "Thinking Introversion." As hypothesized, Neuroticism was related negatively and Extraversion was related positively to most concurrent measures of well-being in both younger and older subsamples. "Thinking Introversion" was related only to positive attitudes toward religion. Predictive correlations between personality and subjective well-being over two-to-ten ($M = 5.3$) and ten-to-seventeen ($M = 12.6$) year intervals confirmed earlier research, and showed that enduring personality dispositions antedate and predict measures of personal adjustment to aging.

Key Words: Well-being, Personality, Longitudinal, Adjustment to aging, Extraversion, Neuroticism.

SUCCESSFUL aging, after Havighurst's pioneering efforts, has been approached from two separate paths. One focuses on the outer, social adjustment defined in terms of the individual's activities and social roles. The other, exemplified in the present study, concerns itself with the inner, subjective experience of personal adjustment. From this second perspective, an individual is considered to have aged successfully if he or she is satisfied with his or her life's accomplishments, retains high morale, or is simply happy. Unsuccessful aging is seen in hypochondriasis, fear of death, a sense of uselessness, loneliness, and depression (Butler & Lewis, 1977). These latter attributes and affects could be considered signs of poor adjustment at any age, but they are particularly relevant to adjustment in elderly adults. As Erikson (1950) pointed out in his formulation of the last stage of psychosocial development, ego integrity vs despair, the major life task of the elderly person is subjective reconciliation and integration. "Adjustment" for the aged cannot be defined by successful socialization, or by productivity, or the establishment of loving interpersonal relationships, since the time for these tasks has passed, according to Erikson. Subjective well-being seems to be the most appropriate criterion for personal adjustment to aging, and much of the literature on gerontology has been devoted to the conceptualization, measurement, and prediction of well-being.

Erikson felt that the likelihood of success at any life task was increased by success at earlier developmental tasks, and thus a mature and well-integrated personality gave the best assurance of successful aging. On purely empirical grounds, a consensus has begun to emerge that enduring personality traits play a major role in successful adaptation to aging (George, 1978; Neugarten, 1977), although the specific aspects of personality responsible for high levels of subjective well-being have not been identified clearly. Similarly, although many personality traits are known to be stable throughout the adult life span (Costa et al., 1980), there is little direct evidence that these traits will predict subjective well-being over a period of years. However, in a recent paper (Costa & McCrae, 1980) using the Bradburn Affect Balance Scale as the criterion, it was found that measures of neuroticism and extraversion derived from the 16PF earlier were significant predictors of subjective well-being 10 years later.

---

[1] Gerontology Research Center (Baltimore), NIA, Dept. of Health and Human Services, Bethesda and the Baltimore City Hospitals, Baltimore, MD 21224. Requests for reprints should be sent to Paul T. Costa, Jr., PhD, Chief, Stress & Coping Section, NIA/Gerontology Research Ctr., Baltimore, MD 21224. Thanks are due to Dr. Clyde E. Martin for his part in collecting the data.

Psychological well-being as measured by the Affect Balance Scale may or may not be equivalent to the subjective well-being identified here as a criterion of successful aging. Some writers have argued that there are important conceptual distinctions to be made between such constructs as happiness, morale, and life satisfaction (Campbell et al., 1976; George, 1979; Lawton, 1977). However, researchers (e.g., Andrews & Withey, 1976) have demonstrated that a wide variety of measures of these constructs show substantial intercorrelation. The present research is based on the empirically guided premise that measures of all these concepts, while not necessarily equivalent, share a common core, and "personal adjustment to aging" is considered within this domain. The success with which previous results relating happiness to personality can be replicated here will speak to the utility of this viewpoint.

The Chicago Attitude Inventory (CAI; Cavan et al., 1949) is used here as a measure of subjective well-being tailored for the study of personal adjustment among the aged. (Analyses of the age relations and longitudinal stability of the CAI scales will be reported in a subsequent manuscript.) The CAI, which was the parent instrument for the widely-used LSI (Neugarten et al., 1961) gives reasonable coverage to most of the important aspects of personal adjustment. Lawton (1977) comments that the CAI "contains a 'Happiness' section that fits clearly into the concepts most central to morale," and Maddox (1963-64) used the CAI as a measure of morale. When supplemented by two items concerning assessment of life and satisfaction with accomplishments the CAI would seem to provide a useful approximation to an adequate operationalization of subjectively defined "successful aging" (Cavan et al., 1949). The CAI was employed in this study because extensive longitudinal data on it were available for analysis. More sophisticated scales have subsequently been developed and should be used in future research (see Lawton, 1977).

DETERMINANTS OF SUBJECTIVE WELL-BEING

It is known that socio-demographic variables are rather weak predictors of subjective well-being or happiness (Andrews & Withey, 1976). Rich or poor, young or old, black or white, most people adjust to their life circumstances. Adaptation-level theory has been invoked to explain this phenomenon (Brickman et al., 1978). But in itself, adaptation theory would predict a "neutral" level of satisfaction or happiness for all people; it could not explain the fact that some people are chronically unhappy while others are characteristically cheerful. Longitudinal studies of well-being (e.g., Palmore & Kivett, 1977) demonstrate that there is some temporal stability in well-being, reflected in stability coefficients of .4 to .5 over periods up to 2 years. Despite circumstances, some individuals seem to be happy people, some unhappy people. We might say that individuals have predispositions to be happy or unhappy, and this suggests that personality traits may be largely responsible for individual differences in happiness, and, by extension, personal adjustment to aging.

Previous research has shown an association between personality variables and happiness (Wilson, 1967). A recent interpretation of this literature (Costa & McCrae, 1980) argued that two sets of traits, which can be identified as facets of the domains of neuroticism and extraversion, are related to psychological well-being. Using measures of these dimensions of personality taken from three standard personality inventories, concurrent correlations were found with the Affect Balance Scale (Bradburn, 1969) and with measures of personal security, hopelessness, and a specially constructed life satisfaction index covering satisfaction in nine areas of life. Ten-year predictive relations between personality measures and the Affect Balance scores were also significant.

The present study uses a different measure — the Guilford-Zimmerman Temperament Survey (GZTS; Guilford et al., 1976) — to operationalize personality, and the CAI to operationalize subjective well-being as personal adjustment to aging. The GZTS yields three factors, which can be identified as neuroticism, extraversion and what Guilford calls "Thinking Introversion," a factor which he believes represents meditative thinking or introspective tendencies. It is hypothesized that neuroticism will be related negatively to personal adjustment whereas extraversion will be related positively. It is not expected that "Thinking Introversion" will be related to the CAI measures. Concurrent correlations are presented for men aged 18 to 49 and 50 to 97. Five and 12-year predictive relations between neuroticism and extraversion factors and CAI variables are examined to determine whether stable personality dispositions are predictive of future states of well-being.

METHODS

Subjects. — Participants in the Baltimore Longitudinal Study are a community-dwelling, generally healthy group of male volunteers, 96% white, who have agreed to return for testing at fixed intervals. The majority (80%) work in or are retired from scientific, professional, or managerial positions. Almost all (93%) are high school graduates, and 71% are college graduates; 88% were married. At the time of their first administration of the CAI subjects ranged in age from 17 to 97. As in all longitudinal studies, there has been some attrition. Subjects who remained in the study tended to be psychologically better adjusted than those who dropped out.

Measures. — The eight sections of the Chicago Attitude Inventory each consist of seven items, to which the respondent marked "agree," "disagree," or "?." Items were coded −1 (poor attitude), 0 (?), or +1 (good attitude) and summed to form eight scales. This method of scoring follows the original method selected by Chicago investigators (Cavan et al., 1949). Scales, together with their internal consistency (coefficient alpha), were: Health (.61), Friends (.49), Work (.19), Economic Security (.43), Religion (.72), Usefulness (.41), Happiness (.61) and Family (.53).

The CAI was supplemented by two global items taken from the Activities section of "Your Activities and Attitudes" questionnaire. Item L6, assessment of life, asks the respondent, "As you look back over your life, in general would you call it: very happy; moderately happy; average; or unhappy?" Since less than 1% of the sample checked "unhappy," this response category was combined with "average" in all analyses. Item L8, satisfaction with accomplishments, asks subjects if they feel: "well-satisfied; reasonably satisfied; or dissatisfied" with what they have accomplished in life. A summary variable was created by summing scores from these items and the eight CAI scales.

A principal factor analysis of these 10 variables for all subjects at first administration showed that three factors had eigenvalues above 1.00, with a mean interscale correlation of .22. Of particular interest here is the first unrotated factor, which represents the common variance among the 10 scales. All variables except religion showed loadings above .40; and the largest loading, .78, was found for the happiness scale.

The measures of personality were orthogonal factor scores derived from the Guilford-Zimmerman Temperament Survey (GZTS). The GZTS is a standard factor based personality inventory which has been used widely in clinical application and research for 25 years. Internal consistency of the scales range from .75 to .87, and 1 year retest reliabilities range from .53 to .80 for college students. Evidence on construct validity is given in the GZTS handbook (Guilford et al., 1976). Over 20 factor analyses reported in the handbook concurred in finding three factors. "Emotional Health" is composed of emotional stability, objectivity, friendliness, and personal relations, with a smaller contribution from masculinity. Since this factor is clearly recognizable as a dimension of neuroticism, factor scores and thus correlations have been reflected, and the factor relabeled Neuroticism by the present authors. "Social Activity" or Extraversion is defined by general activity, ascendance, and sociability. "Thinking Introversion" is a doublet defined by restraint and thoughtfulness. Factor scores were derived from an analysis of the GZTS on the present sample which replicated previous factor results (McCrae et al., in press).

Procedure. — Each subject was given the CAI individually with standard instructions and completed the questionnaire during the remainder of his 3-day visit to the Gerontology Research Ctr. The CAI was given on the first or second visit, and readministered on the fifth and ninth visits. Subjects were given the GZTS individually with standard instructions on their first or second visit to the GRC. For each item, subjects chose "yes" or "no" or "?." Each scale consists of 30 items, but only "yes" or "no" responses contribute to the total score. A measure was invalidated for any scale with more than three "?" responses. For subjects missing up to five scales factor scores were estimated from the remaining scales by assigning mean values for the missing scales. Subjects with less than five valid scales were excluded. In all analyses correlations were based on all available cases; because of missing data the specific *N*s vary and are given in the table.

Table 1. Retest Correlations for Two Intervals for 10 Personal Adjustment Variables.

Variables	First to Second Administration[a]		First to Third Administration[b]	
	r	*N*	*r*	*N*
Assessment of Life	.56***	419	.59***	143
Satisfaction with Accomplishments	.49***	422	.46***	144
Health	.48***	366	.20*	134
Friends	.51***	384	.50***	134
Work	.48***	388	.21**	133
Economic Security	.52***	372	.20*	127
Religion	.80***	372	.78***	128
Usefulness	.52***	391	.30***	139
Happiness	.50***	376	.40***	132
Family	.40***	358	.32***	130

[a] Interval = 4 to 8 years, M = 6.3 years.
[b] Interval = 10 to 19 years, M = 12.8 years.
*$p < .05$; **$p < .01$; ***$p < .001$.

Analyses. — In the first analyses, Pearson correlations were calculated between the CAI scales and the contemporaneous GZTS factors. Since these analyses were intended to represent concurrent relations, subjects were selected who had taken the GZTS within 2 years of their first CAI. Analyses were conducted for two subsamples: 418 men aged 18 to 49 (M = 36.6 years) and 391 men age 50 to 97 (M = 64.3 years).

In the second analysis predictive relations between personality and adjustment were sought by correlating GZTS Neuroticism and Extraversion factor scores with scales from the second administration of the CAI for the total sample. The time interval separating these two was from 2 to 10 years, with a mean of 5.3 years. In the third analysis GZTS factor scores from the first administration were correlated with the CAI scales from the third administration to provide long-term predictive coefficients. The time interval separating the two was from 10 to 17 years with a mean of 12.6 years. Under the hypothesis that neuroticism is negatively and extraversion is positively related to well-being, one-tailed significance tests were employed in the second and third analyses.

Table 2. Concurrent Prediction of Personal Adjustment from GZTS Factors for Young (18-49) and Old (50-97).

	Young GZTS Factors				Old GZTS Factors			
Variables	*N*	I[a]	II[b]	III[c]	*N*	I[a]	II[b]	III[c]
Assessment of Life	378	−.21***	.27***	.00	342	−.26***	.15**	−.03
Satisfaction with Accomplishments	380	−.17***	.26***	.06	343	−.15**	.12*	.04
Health	372	−.12*	.25***	−.05	308	−.15**	.23***	.00
Friends	370	−.33***	.22**	.02	313	−.17**	.24***	.04
Work	372	−.19***	.11*	.12*	315	−.06	.20***	.11
Economic Security	371	−.26***	.11*	.01	307	−.16**	.09	.01
Religion	367	−.16**	.07	.14**	307	.04	.08	.18**
Usefulness	376	−.24***	.15***	.05	321	−.13*	.26***	.08
Happiness	370	−.31***	.32***	−.07	313	−.19***	.25***	−.05
Family	360	−.27***	.05	.05	306	−.10	.14*	.01
Sum	320	−.40***	.26***	.05	237	−.18**	.37***	.04

[a] Neuroticism; [b] Extraversion; [c] "Thinking Introversion".
*$p < .05$; **$p < .01$; ***$p < .001$.

RESULTS

Table 1 presents the retest correlations for first to second and first to third administrations. Correlations are moderate except in the case of Religion, which is quite stable. The lower retest coefficients of the remaining variables may indicate either unreliability of measurement or sensitivity to change. As with the relatively low internal consistency, the modest retest reliability may attenuate correlations with personality variables.

As an omnibus test for the relation between GZTS factors and CAI variables, canonical correlations were computed. These were .52, $p < .001$, for the younger group, and .49, $p < .001$, for the older group. Table 2 presents the concurrent correlation coefficients for the GZTS factors of Neuroticism, Extraversion, and "Thinking Introversion" with the CAI variables and their sum for the two age groups. As Table 2 shows, both Neuroticism and Extraversion are correlated significantly with reports of adjustment and positive attitudes for almost all areas of life for both the younger and older age groups. Only positive attitudes toward religion shows a pattern of independence from the first two factors, and it is also the only variable significantly related to "Thinking Introversion" for both age groups.

Canonical correlations between the first two GZTS factors and the ten CAI variables were calculated for the two predictive intervals as an omnibus test. For the two to ten year interval, the correlation of .46 was significant at the .001 level. Due to small N, the correlation for the 10 to 17 year interval, also .46, fell just short of conventional significance levels, $p = .054$. However, under the prior hypotheses advanced, a one-tailed test would be more appropriate, and would reach significance. Table 3 gives the predictive relations between Neuroticism and Extraversion factors and the CAI variables. Over an interval of from two to ten years, all measures of personal adjustment except religion are predicted by low Neuroticism factor scores, and all but economic security are predicted by Extraversion scores. Over an interval of from 10 to 17 years, Neuroticism significantly predicts feelings of usefulness, happiness, and positive attitudes toward family; and extraversion significantly predicts these three attitudes, the friends scale, and the single items of assessment of life and satisfaction with accomplishments. Although statistically significant and consistent with theory, these correlations are quite small. This may be due to the increasing selectivity of the third administration sample as well as to the long predictive interval. In addition, it may be due to the unreliability of the CAI scales. Using the more reliable sum of these scales, both factors show somewhat larger correlations of −.25 and .27.

Both Neuroticism and Extraversion are related to positive attitudes toward life in both age groups, and both show predictive relations for periods up to 17 years. In this sample, Extraversion appears to be a somewhat

Table 3. Prediction of Personal Adjustment Variables From GZTS Neuroticism and Extraversion Over Two Intervals.

	2 to 10 Year Prediction			10 to 17 Year Prediction		
Variables	*N*	Neuroticism	Extraversion	*N*	Neuroticism	Extraversion
Assessment of Life	422	−.19***	.25***	116	−.08	.35***
Satisfaction with Accomplishments	423	−.19***	.12**	117	−.11	.26**
Health	385	−.17***	.17***	114	−.03	−.03
Friends	404	−.17***	.13**	114	−.15	.22*
Work	404	−.09	.15**	113	−.14	.05
Economic Security	394	−.22***	.00	113	−.06	.04
Religion	392	−.03	.09	111	−.09	.13
Usefulness	401	−.13**	.24***	116	−.17*	.23**
Happiness	396	−.17***	.20***	114	−.17*	.24**
Family	381	−.16***	.18***	111	−.18*	.31***
Sum	324	−.27***	.28***	97	−.25**	.27**

*p (one-tailed) < .05; **p (one-tailed) < .01; ***p (one-tailed) < .001.

better predictor than Neuroticism. This may be because Neuroticism is particularly relevant to the prediction of the "unhappy" portion of the affect balance dimension, and so small a proportion of the present sample is unhappy. Here the predictive task may be to distinguish the "moderately happy" from the "very happy", and Extraversion appears more effective in making this distinction (Costa & McCrae, 1980).

Only religion seems unrelated to Extraversion and Neuroticism, being related instead to "Thinking Introversion." On *a priori* grounds, it could be argued that the proreligious attitudes measured by the CAI religion section should not be regarded as a component of personal adjustment (although *satisfaction* with one's religious faith or lack of it might be). Havighurst and Albrecht (1953) report that religion alone does not seem to intercorrelate with the other attitude sections, and a similar pattern was observed in the present data. By contrast, the happiness section showed the highest mean correlation with other sections, consistent with the premise that there is a general dimension of well-being in all but the religion sections.

DISCUSSION

Theorists have drawn attention to the host of distinct influences which are presumed to operate in determining levels and varieties of well-being. Thus, it may not be the absolute level of attainment that is responsible for subjective satisfaction or dissatisfaction, but the discrepancy between expectations and achievements. The distinction between cognitive judgements of satisfaction and affective reactions is thought by some to be significant (Campbell et al., 1976). Important differences in content between measures of morale, well-being, and life satisfaction have also been noted (George, 1979; Lawton, 1977).

Although these distinctions may be important in certain contexts, the burden of proof remains on those who advocate them. Whereas studies showing moderately high levels of intercorrelation give evidence of convergent validity for many measures in this area, there is no convincing evidence of discriminant validity for any of them. In general, regardless of their theoretical rationales, measures of well-being, morale, or life satisfaction tend to behave in the same ways.

One of the important ways in which they behave similarly is in their relation to personality variables. The present findings argue that there is a common core to measures of morale, well-being, or personal adjustment to aging. For simplicity, we can call this core element "subjective well-being." However defined or measured, individuals higher in subjective well-being appear to be lower in neuroticism and higher in extraversion.

Research implications. — Considerable research has been devoted to efforts to understand the determinants of life satisfaction, morale, or well-being in the elderly. Social interaction has often been identified as a correlate of life satisfaction, but Conner et al. (1979) conclude that quantitative variables like scope or frequency of interaction are poor predictors of well-being and that "it is in the quality of the interactional experience that a broader understanding of adjustment to the process of aging will ultimately be found." One interpretation of the association of extraversion and neuroticism with well-being fits this prescription for research on the quality of interactional experience. Extraverts characteristically show warm, expressive interactions, while the personal relations of individuals high in neuroticism are often characterized by hostility or self-consciousness. Indeed, a questionnaire designed to assess the quality of interactional experiences might well yield factors identifiable as neuroticism and extraversion. More direct tests of this position might involve observer ratings of interactions in conjunction with self-reports of personality and well-being.

It is not particularly surprising that individuals who are dispositionally anxious, hostile, or depressed should be more unhappy than others; or that those who are warm, friendly, or active should be happier. But the implications of this fact have seldom been recognized. The domains of neuroticism and extraversion have a host of behavioral correlates, and one or the other of these personality dimensions may be the "third variable" which accounts substantially for correlations between morale or life satisfaction and such variables as perceived health, marital status, activity level, or the use of a confidant.

Consider the relation between well-being and self-reported health. Larson (1978) concludes that perceived health is the variable most highly correlated with psychological well-being. An extensive literature has demonstrated that perceived health is a function of both objective health status (Maddox & Douglas, 1973) and neuroticism (Blazer & Houpt, 1979). It is possible that the observed correlation between well-being and perceived health reflects the influence of neuroticism more than that of objective health. If so, interventions which improve the health of elderly adults would not be expected to produce a long-term improvement in their psychological well-being. Adjusted extraverts are likely to be cheerful even when sick, and neurotic introverts will find a new source of unhappiness when health is no longer a problem. One implication of this interpretation is that research in this area should take steps to control for personality dispositions. Independent, objective measures of health status could be used in place of self ratings of health. Pre- and post-intervention designs could be used in which each subject serves as his/her own control in regard to personality dispositions. Most simply, standard measures of neuroticism and extraversion could be administered in correlational studies and partial correlations controlling for scores on these measures could be reported. Similar arguments would apply to research on other correlates of psychological well-being.

Finally, intervention studies should also take into account the principles of adaptation level and individual differences in selecting outcome criteria. Not only is morale an inappropriate criterion for many kinds of interventions (Rosow, 1977), it is also insensitive to changes which may make appreciable differences in the quality of life. As Carp (1977) recommends, "specific program-relevant responses" should be substituted for global measures of well-being.

REFERENCES

Andrews, F. M., & Withey, S. B. *Social indicators of well-being: Americans' perceptions of life quality.* Plenum, New York, 1976.

Blazer, D. G., & Houpt, J. L. Perception of poor health in the healthy older adult. *Journal of the American Geriatric Society*, 1979, *27*, 330-334.

Bradburn, N. M. *The structure of psychological well-being.* Aldine, Chicago, 1969.

Brickman, P., Coates, D., & Janoff-Bulman, R. Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology*, 1978, *36*, 917-927.

Butler, R. N., & Lewis, M. I. *Aging and mental health: Positive psychosocial approaches*, (2nd ed.) C. V. Mosby, St. Louis, 1977.

Campbell, A., Converse, P. E., & Rodgers, W. L. *The quality of American life: Perceptions, evaluations, and satisfactions.* Russell Sage Fdn., New York, 1976.

Carp, F. M. Morale: What questions are we asking of whom? In C. N. Nydegger (Ed.), *Measuring morale: A guide to effective assessment.* The Gerontological Society, Washington, DC, 1977.

Cavan, R. S., Burgess, E. W., Havighurst, R. J., & Goldhamer, H. *Personal adjustment in old age.* Science Research, Chicago, 1949.

Connor, K. A., Powers, E. A., & Bultena, G. L. Social interaction and life satisfaction: An assessment of late-life patterns. *Journal of Gerontology*, 1979, *34*, 116-121.

Costa, P. T., Jr., & McCrae, R. R. The influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, 1980, *38*, 668-678.

Costa, P. T., Jr., McCrae, R. R., & Arenberg, D. Enduring dispositions in adult males. *Journal of Personality and Social Psychology*, 1980, *38*, 793-800.

Erikson, E. *Childhood and society.* W. W. Norton, New York, 1950.

George, L. K. The impact of personality and social status factors upon levels of activity and psychological well-being. *Journal of Gerontology*, 1978, *33*, 840-847.

George, L. K. The happiness syndrome: Methodological and substantive issues in the study of social psychological well-being in adulthood. *The Gerontologist*, 1979, *19*, 210-216.

Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. *The Guilford-Zimmerman temperament survey handbook: Twenty-five years of research and application.* Robert R. Knapp, San Diego, 1976.

Havighurst, R. J., & Albrecht, R. *Older people.* Longmans, Green, New York, 1953.

Larson, R. Thirty years of research on the subjective well-being of older Americans. *Journal of Gerontology*, 1978, *33*, 109-125.

Lawton, M. P. Morale: What are we measuring? In C. N. Nydegger (Ed.), *Measuring morale: A guide to effective assessment.* The Gerontological Society, Washington, DC, 1977.

Maddox, G. L. Activity and morale: A longitudinal study of selected elderly subjects. *Social Forces*, 1963-64, *42*, 195-204.

Maddox, G. L., & Douglas, E. G. Self-assessment of health: A longitudinal study of elderly subjects. *Journal of Health and Social Behavior*, 1973, *14*, 87-93.

McCrae, R. R., Costa, P. T., Jr., & Arenberg, D. Constancy of adult personality structure in males: Longitudinal, cross-sectional and times-of-measurement analyses. *Journal of Gerontology*, in press.

Neugarten, B. L. Personality and aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging.* Van Nostrand Reinhold, New York, 1977.

Neugarten, B. L., Havighurst, R. J., & Tobin, S. S. Measurement of life satisfaction. *Journal of Geron-*

tology, 1961, *16*, 134-143.

Palmore, E., & Kivett, V. Change in life satisfaction: A longitudinal study of persons aged 46-70. *Journal of Gerontology*, 1977, *32*, 311-316.

Rosow, I. Morale: Concept and measurement. In C. N. Nydegger (Ed.), *Measuring morale: A guide to effective assessment*. The Gerontological Society, Washington, DC, 1977.

Wilson, W. Correlates of avowed happiness. *Psychological Bulletin*, 1967, *67* 294-306.

Elias, M.F.; and Streeten, D.H.P. (Eds.)
Hypertension and Cognitive Processes.
Mt. Desert, Me.: Beech Hill, 1980

New Longitudinal Data on the Question of Whether Hypertension Influences Intellectual Performance

Paul T. Costa, Jr., Ph.D.,
and
Nathan W. Shock, Ph.D.

An extensive literature [see 2] on aging and cognitive functioning has documented that declines in performance are to be expected in a number of areas, including speed of reaction and short-term memory. Because elevated blood pressure is a frequent accompaniment of age, it has been suggested that these cognitive declines may be caused at least in part by hypertension.

The evidence on which this hypothesis must be evaluated is very mixed. Most studies [4; 5; 9; 11; 10; 12] have found poorer performance on a variety of tasks among hypertensives, but some studies [6; 7] have found improved performance associated with elevated blood pressure. Equally puzzling is the relation between age, hypertension, and intellectual performance. Using the WAIS, Schultz and his colleagues [10] found that hypertension was detrimental for young but not middle-aged men and women, whereas Wilkie and Eisdorfer [12] reported cross-sectional findings that showed blood pressure was negatively correlated with intelligence for 70-year-olds, but not for 60-year-olds. These complex relationships are difficult to interpret.

In part, these anomalies are due to the complexity of the variables under investigation. As recent researchers have emphasized, [e.g. 4], hypertension is not a single clinical entity but is of various kinds, and these etiological differences may influence any relations with cognition. Most research has been done on clinic populations undergoing treatment, and the effects of different kinds and levels of medication are not fully understood. On the cognitive side, the particular tests employed may also be expected to show different effects. Vocabulary, for example, generally does not decline with age, and may even increase. Visual memory, by contrast, decreases dramatically as Arenberg [1] has shown. Since different tests incorporate different proportions of these "crystalized" and "fluid" components of intellect, comparison is difficult.

Small sample sizes may also contribute to the inconsistency of the effects of blood pressure on intellectual performance and decline. The effects of hypertension (if there are any) are small in comparison with the range of individual differences in intelligence, and small effects are often obscured when small samples are used. For example, Wilkie and Eisdorfer's failure to find significant differences in the decline for 70-79 year old borderlines from that of normotensives may be due to the small *N* (i.e., 8) in the borderline group. Again, although some investigators have carefully screened and matched subjects [e.g., 4; Chapter 3, this volume], sampling procedures are sometimes less than optimal, since the available subjects may be the self-selected patients of hypertension clinics. Sample fluctuations may account for some of the variability in reported findings.

Finally, investigators [5] have pointed out that longitudinal studies are required if the long term effects of hypertension on performance are to be assessed. Longitudinal studies, which provide serial observations on the same individuals over time, permit determination of antecedent-consequent relations.

Table 1

Time Limits and Number of Items for Army Alpha Subtests

Subtest	Time Limit (Minutes)	Number of Items (Possible Score)
Following Directions	2.25	12
Arithmetic Problems	5.00	20
Practical Judgment	1.50	16
Synonym-Antonym	1.50	40
Disarranged Sentences	2.00	24
Number Series Completion	3.00	20
Analogies	3.00	40
General Information	4.00	40

The present study uses the Army Alpha Test, a general intelligence test with both "fluid" and "crystallized" components. Relatively large samples ($N = 350$ and $N = 51$) are used, and basic information on health status and medication is available. Finally, longitudinal follow-ups over a period of up to eight years allow a test of the long-term effects of hypertension on these cognitive tasks.

The Army Alpha has been given to BLSA participants over a twenty-year period. The test is administered individually under standard timed conditions. After scores are determined, the test is returned to the subject and he is given an opportunity to finish the test at leisure. Two sets of scores are thus obtained: a timed set, where speed may be important, and an untimed set, where speed should not be a factor. Unfortunately, about 5 percent of the subjects, mostly older men, declined to complete the untimed test; consequently, complete data are not available for them.

All subjects were given Form A of the Alpha [3] on their first visit. Thereafter, subjects were treated differently according to age. On their fifth visit to the GRC, after an interval of from 4-8 years, subjects under age 70 were given a parallel Form B of the Army Alpha, in order to avoid practice effects. Subjects who are over 70 are given re-administrations of Form A at each annual visit. Thus, two sets of longitudinal data are available for analysis: (1) Re-test data after four-to-eight years on Form B for 350 men initially aged 17-65, and (2) data from six administrations of Form A spanning five to eleven years for 51 men initially aged 66-84.

BLOOD PRESSURE AND INTELLIGENCE AMONG MEN UNDER 65

An extensive series of analyses were carried out on the two-point data for men under 65. In the first set of analyses, subjects were classified into three age groups: 20-39, 40-49, and 50-65. They were cross-classified as low, average, or high in blood pressure on the basis of three sources of data. The first was the average basal blood pressure at the first visit described elsewhere (see Chapter 7). Average systolic and diastolic values were trichotomized for the entire group to give roughly equal thirds. This will be referred to as the Basal classification. The next data source used was an average of right and left casual sitting pressures taken by a physician at the subject's first visit. For this First Casual classification, subjects were classed as "low" if their systolic pressure was below 120, or their diastolic pressure was below 80. Subjects were classified as "high" if their systolic pressure was above 140 or their diastolic pressure was above 90. The remaining subjects were classified in a "middle" group. Finally, since there is concern over the possibility of a "shock effect" on the first visit, averaged right and left casual blood pressures from the second visit were also employed, using the same cut-off points as for the First Casual classification.

This procedure of using three different bases for classifying hypertension groups (Basal, First Casual, and Second Casual) complicates the analyses, and increases the probability of accepting a chance finding as a true effect. On the other hand, it allows for a kind of replication of findings across three independent operationalizations of blood pressure categories. If elevated blood pressure (within the range represented in this sample) exerts a real influence on cognitive functioning, it ought to be seen using any reasonable basis of analysis. Similarly, we would expect that classifications based on systolic and diastolic pressures would yield similar results, although it is certainly possible that one or the other is more predictive of cognitive decline.

In the first analyses, then, each of the eight timed subtests and the total Alpha score were examined for six (systolic or diastolic by Basal, First Casual or Second Casual) definitions of blood pressure classification. Age group and time were also used as a classifying variable. Fifty-four F ratios were computed in which blood pressure was a main effect, and 162 F ratios in which it was a part of an interaction term. One main effect was statistically significant ($p<.05$). There was no significant interaction with age group, although there was one significant ($p<.05$) interaction with time and six significant ($p<.05$) age by time by blood pressure interactions. One of these was found for the Second Casual pressure on both systolic and diastolic pressures; another was found for both First and Second Casual but not Basal pressures. The other interactions were not replicated. By chance, we would expect 5% (or 8) of the 162 interaction terms to be significant at the .05 level. We found seven. Clearly, there is no evidence in these data for any effect of hypertension (within the present BP ranges) on Army Alpha performance among men under 65.

A similar set of analyses were conducted on the seven sub-tests of the Alpha which were also given under untimed conditions, and the untimed total. (The first subtest, "Following Directions", requires verbal instructions from an examiner, so it is not included among the untimed tests.) Here, three of 48 main effects, and five of 144 interaction terms were significant. Again, chance seems to be a compelling explanation.

However, all of these analyses were undertaken on an *unscreened* population, in which the influence of chronic illness or drugs, particularly antihypertensive drugs, may obscure a true effect. Consequently, additional analyses were performed in which subjects taking medication or diagnosed as having a serious chronic illness (other than hypertension) at any point during their first five visits to the GRC were excluded. This left, of course, a much smaller sample with 117 men, relatively few of them in the "high" hypertension group. Since there was no evidence of age by hypertension interactions, subjects were collapsed across age groups, and age was used as a covariate in these analyses.

Using timed subtests and totals as dependent variables, 13 significant main effects were observed, and five interactions. In addition, some of the main effects were replicated. Thus the "high" group performed least well on the "Following Direction" and "Number Series" subtests based on the Basal diastolic, the First Casual diastolic, and the First Casual systolic classifications. These effects are weak and by no means consistently replicated, but they give at least mixed evidence that higher blood pressure may be detrimental to some kinds of cognitive performances among otherwise healthy men.

Somewhat stronger evidence was found when the untimed condition of the Alpha was analyzed. Replicated main effects were seen for "Arithmetic Problems", "Synonym-Antonym", "Disarranged Sentences", and "Number Series Completion", as well as for the Total untimed score. In all of these, the "high" group showed the poorest performance. Again, however, none of the effects was strong, and no finding was statistically significant in all six possible replications.

In addition, it is possible that the apparent effects of elevated blood pressure are due to the association of both variables with age. By using age as a covariate, statistical adjustments have been made for the *linear* effects of age. But some cognitive performances show a curvilinear, accelerated decline with age which is not fully eliminated in these analyses.

Finally, it should be noted that there is no replicated blood pressure group by time interaction. Over a period of from four to eight years, moderately hypertensive individuals do not decline in performance any more than normotensives. This is a particularly important negative finding. If hypertension *causes* a decline in cognitive capacity, it might be expected that chronic hypertension should result in *progressive* impairment of cognitive performance. The fact that this is not seen argues for the hypothesis that association between hypertension and cognition is not directly causal. Interestingly, parallel results have been found in mice [4]. Occupational, dietary, constitutional, or other factors related to both intelligence and blood pressure levels may account for the observed association.

BLOOD PRESSURE AND INTELLIGENCE AMONG ELDERLY MEN

Wilkie and Eisdorfer [12] reported different patterns of association between WAIS scores and blood pressure for 60-69 year olds and 70-79 year olds. In the 60 year olds, there was no initial correlation between blood pressure and WAIS scores. Over a ten year interval, hypertensives declined in performance while normotensives did not change and borderline subjects significantly increased. Among the 70 year olds, initial levels of blood pressure were negatively correlated with WAIS scores, and over a ten year interval both normotensives and borderlines declined, though not differentially. It is important to note that none of the oldest hypertensives returned for the ten year retest. The notable finding here is that borderline hypertensives (diastolic pressures of 96-105) showed increased performance among 60 year olds and decreased performance among 70 year olds.

In an attempt to more clearly parallel this study with BLSA data, 51 men were identified who had taken the Army Alpha, Form A, on six occasions. They ranged in age from 66 to 84 at first administration and from 74 to 93 at sixth administration. Similar to subjects in the Wilkie and Eisdorfer study, all were free of definite cerebrovascular disease. Physicians' Second Casual blood pressures, which were used as the basis for analyses, ranged from 98 to 186 for systolic and 55 to 115 for diastolic. Since a number of subjects had not completed the Alpha under untimed conditions, only results from the timed administration are presented.

In a first analysis, all subjects were classified as (1) free of drugs and major chronic diseases (other than hypertension) or (2) other. No significant differences were observed on any of the eight subtests or total scores. It was therefore possible to collapse across these two categories in all subsequent analyses. Repeated measures analyses of variance were performed on the Alpha subtests and total score using six levels of the repeated factor, and classifying subjects by age and blood pressure. Two levels of the classifying factors were used: above and below 72.5 years at initial testing, mean age = 69.7 ($N = 28$) versus 75.2 ($N = 23$); and above or below systolic pressure of 140, mean systolic = 118.7 ($N = 27$) versus 154.7 ($N = 24$); or diastolic pressure of 90, mean diastolic = 75.6 ($N = 37$) versus 96.4 ($N = 14$). The reader should note that the hypertensive subjects in our study (as in the studies of Elias and colleagues) more closely resemble Wilkie and Eisdorfer's [12] *borderline* hypertensives than their hypertensive group.

On only one of the subtests, "Practical Judgment", was there a significant main effect for blood pressure classification. Hypertensive men were somewhat lower performers than normotensive men in this task, although only the classification based on systolic pressure showed this effect. Signifi-

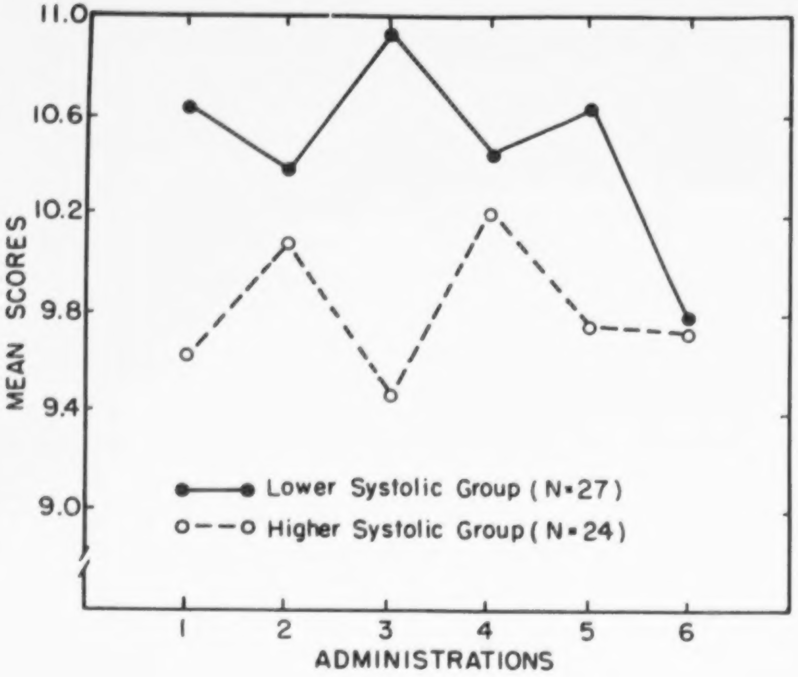


Figure 1. Mean "Number Series Completion" scores for lower and higher systolic pressure groups at six administrations, collapsed across age groups.

cant ($p<.05$) two-way interactions were observed between time and both systolic and diastolic pressure groups on "Number Series Completion". It might be expected that normotensives would remain relatively constant over the six administrations or show only a modest decline while hypertensives would show a more progressive and precipitous decline. However, Figure 1, which shows "Number Series" scores for systolic blood pressure groups, provides little support for this hypothesis. The saw-tooth patterns in this figure probably reflect errors of measurement or chance variation more than the effects of blood pressure. Similar results are seen for diastolic groups.

The Wilkie and Eisdorfer study [12] might suggest that different patterns would be seen for younger and lower blood pressure groups. This would imply a triple interaction between time, blood pressure groups, and age groups. No such interactions were significant. As in the study of men under 65, men over 65 do not appear to change in cognitive performance as a function of elevated blood pressures.

Indeed, it is rather remarkable how well all these men preserve their intellectual capacities as measured by the Alpha. One subject, for example, entered the study at age 71 with a Second Casual systolic reading of 135 and diastolic reading of 85. During the next six years he obtained scores of 186, 188, 194, 193, 190, and 184 on his total timed Army Alpha. These scores are above the 90th percentile based on norms for college freshmen! In many of our older subjects, neither age nor hypertension seems to impair intelligence seriously.

CAUTIONS AND LIMITATIONS

The two studies reported here both suggest that elevated blood pressure levels have little if any effect on intellectual performance on the Army Alpha for men in the age range from 17 to 88.

These conclusions, however, must be tempered by a number of qualifications and limitations in the present study. Most importantly, the level of elevation of blood pressure seen in the present group is relatively modest: unlike the Wilkie and Eisdorfer subjects there are no cases of untreated, severe hypertension. Although Schultz and colleagues found an effect on younger subjects (but not lower) for moderate elevations of blood pressure, it could be argued that the deleterious effects of hypertension on cognition generally manifest themselves only in extreme cases. The present results do not address the effects of extreme hypertension, but they do speak to the much more common problems of borderline and moderate hypertension. Recent research has clearly indicated that even small chronic elevations of pressure result in an increased risk of mortality from coronary heart disease, but a similar claim does not seem to be justified with regard to the danger of intellectual impairment.

The present sample is also atypical in other respects. The level of education is far higher than that of the general population. It has often been speculated that continued use of intellectual functions may delay their age-related decline. Similarly, exercise of cognitive skills may retard or prevent the adverse effects of hypertension. If that were the case, the negative findings of the present study would not preclude the possibility that even moderate levels of hypertension might be a cause of intellectual decline among the less gifted.

This discussion has been predicated on the widely held assumption that the effects of hypertension on mental performance are necessarily adverse. The mechanisms, if any, by which elevated blood pressures lead to poorer performance have not been demonstrated, but it may be supposed that they are similar to those involved in stroke. By this *hypothesis,* hypertensives are "brain damaged", at least to some degree. Certainly hypertension is one of the recognized risk factors for stroke, and certainly stroke can have devastating consequences for mental functioning. But it has not yet been demonstrated that uncomplicated essential hypertension has any effect at all.

It has also been suggested [12] that elevated blood pressure may have beneficial cognitive effects for the elderly. According to this view, generalized cerebral arteriosclerosis, an almost universal process among the elderly, reduces blood flow to the brain. Increased blood pressure, however, might be able to increase the blood flow and in turn the supply of oxygen to the brain, and thus help to sustain intellectual performance. However, Marsden [8], citing more recent evidence, has stated that the "human cerebral circulation has a remarkable capacity to adjust blood flow in the face of wide fluctuations in perfusion pressure...As a result, even severe reductions in the size of the lumen of major blood vessels supplying the brain, and large drops in systemic blood pressure, do not compromise flow (p. 103)."

At present, evidence from a number of sources points to the conclusion that uncomplicated essential hypertension is more likely to be detrimental than benefical if it has any direct effect on cognitive capacity at all. There are sound medical grounds for attempting to reduce hypertension, and nothing in the cognitive performance literature would contest that approach.

SUMMARY

Two multivariate, longitudinal investigations of the effects of elevated blood pressure on intellectual performance were reported. Sub-scale and total scores from the Army Alpha test, Forms A and B, were used as dependent measures. In the first study subjects were 350 men aged 17-65 retested after 4-8 years. Repeated measures analyses of variance with three levels of age and three kinds of blood pressure were conducted, using three different sets of blood pressure recordings to operationalize blood pressure groups. No reliable effects for either factor were noted. When analyses were restricted to 117 men free from drug therapy and major illnesses other than hypertension, only a few effects were replicated across different operationalizations of blood pressure groups, and showed poorer performance for individuals with higher levels of blood pressure. Unlike the findings of Schultz et al. [10], no age-by-blood pressure interactions were observed. There was no evidence that hypertensive individuals declined more in intellectual performance than normotensives over the retest interval.

In the second analysis, 51 men aged 66-84 retested on five occasions within an 11 year interval were classified into two levels of blood pressure and two levels of age. Only one sub-test showed significantly poorer performance for hypertensive subjects. Over the six time points of this study, there was no evidence of monotonic decline among hypertensives relative to normotensives.

These results contradict the hypothesis that moderate elevations of blood pressure are beneficial for continued intellectual performance among older subjects. However, they also provide little or no support for the hypothesis that hypertension is associated with a decline in intellectual capacity and performance.

REFERENCES

1. Arenberg, D. Differences and changes with age in the Benton Visual Retension Test. *Journal of Gerontology*, 1978, *33*, 534-540.

2. Botwinick, J. Intellectual abilities. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold Co., 1977.

3. Bregman, E. O. *Manual - revisions of the Army Alpha examinations, Form A and Form B.* New York: The Psychological Corp., 1947.

4. Elias, M. F. Some contributions of genetic selection to the study of hypertension and behavior over the life span: Methodologic considerations and useful future directions. *Birth Defects: Original Article Series*, 1978, *14*, 121-156.

5. Elias, M. F. The relationship of hypertension to cognitive functioning. In D. G. Stein (Ed.), *The psychobiology of aging: Proglems and perspectives.* Amsterdam: Elsevier, in press.

6. Hertzog, C., Schaie, K. W., & Gribbin, K. Cardiovascular diseases and change in intellectual functioning from middle to old age. *Journal of Gerontology*, 1978, *33*, 872-883.

7. Karasawa, A., Kawashima, K., & Kasahara, H. Mental aging and its medico-psycho-social background in very old Japanese. *Journal of Gerontology*, 1979, *34*, 680-686.

8. Marsden, C. D. The diagnosis of dementia. In A. D. Issacs & F. Post (Eds.), *Studies in geriatric psychiatry*. New York: Wiley, 1978.

9. Pentz, C. A., III, Elias, M. F., Wood, W. G., Schultz, N. R., & Dineen, J. Relationship of age and essential hypertention to neuropsychological test performance. *Experimental Aging Research*, 1979, *5*, 351-372.

10. Schultz, N. R., Dineen, J. T., Elias, M. F., Pentz, C. A., & Wood, W. G. WAIS performance for different groups of hypertensive and control subjects during the administration of a diuretic. *Journal of Gerontology*, 1979, *34*, 246-253.

11. Spieth, W. Cardiovascular health status, age, and psychological performance. *Journal of Gerontology*, 1979, *34*, 246-253.

12. Wilkie, F. L., & Eisdorfer, C. Intelligence and blood pressure in the aged. *Science*, 1971, *172*, 959-962.

Journal of Gerontology
1978, Vol. 33, No. 5, 737-747

Age Changes, Cohort Differences, and Cultural Change on the Guilford-Zimmerman Temperament Survey[1]

Karen Douglas, BA, and David Arenberg, PhD[2]

The relation between adult age and temperament was investigated using the Guilford-Zimmerman Temperament Survey. Between the years 1958 and 1974, the GZTS was administered to 915 men from 17 to 98 years of age in the Baltimore Longitudinal Study. Repeated measures were obtained for 336 men approximately 7 years after initial testing. Each GZTS scale was analyzed cross-sectionally and longitudinally. Sequential analyses of independent samples were also carried out in an attempt to separate the effects of maturation from those of generational and cultural change. Results showed that Masculinity declined at all ages, but General Activity declined only after age 50. Thoughtfulness and Personal Relations showed cultural declines during the time period studied, whereas Friendliness showed a long-term cultural decline. Later-born cohorts were lower in Restraint and higher in Ascendance than early-born cohorts.

THE purpose of this study was to investigate the relationship between age and the ten scales of the Guilford-Zimmerman Temperament Survey (GZTS). In 1958, when data collection for the GZTS began in the Baltimore Longitudinal Study[3], few cross-sectional studies of objective personality scales and adult age had been reported; the only longitudinal study of adults (Kelly, 1955) was restricted to the period from young adult to middle age. A number of longitudinal studies have been reported since 1958 (see Neugarten, 1977 for a review), but few studies (Schaie & Parham, 1976; Woodruff & Birren, 1972) have attempted to extricate maturational from cultural effects. Such an attempt was made in the current study by including cross-sectional, conventional longitudinal, and sequential designs (see Baltes, 1968; Schaie, 1965, 1970). Due to the lack of well-developed theories of personality and aging, the study was not designed to investigate specific hypotheses; rather, the intent was to provide descriptive data which would serve as a foundation for future theory formulation (see Neugarten, 1977).

The GZTS is a questionnaire which provides an assessment of ten traits: General Activity, Restraint, Ascendance, Sociability, Emotional Stability, Objectivity, Friendliness, Thoughtfulness, Personal Relations, and Masculinity. Three cross-sectional studies investigated age differences in men on the GZTS scales (Bendig, 1960; Titus & Goss, 1969; Wagner, 1960), but the results were inconsistent. In all three studies Ascendance decreased with age. In two of the three studies, Sociability decreased with age, and Restraint and Personal Relations increased. General Activity, Masculinity, Emotional Stability, and Objectivity each decreased with age in one of the three studies, and Friendliness and Thoughtfulness were not age related in any study. For a more detailed review of the GZTS, see Guilford et al., (1976).

The four types of analyses reported in this paper addressed the following questions: (a) Over a wide range of adult age, are there age differences in the GZTS, and how do these differences compare with those previously reported? Differences could be due to maturational change or cohort differences. (b) Are there age changes within individuals on GZTS scale measures, and for which scales is the magnitude of change related to age? Longitudinal changes could be due to maturation, cultural change during the interval between measures, selective attrition, or even retest effects (repeating the measure). (c) Are the age changes in repeated measures of the same subjects confirmed in comparisons of independent samples of the same cohort measured at different times? Under ideal sampling conditions, such confirmation eliminates retest effects and some effects of selective attrition as factors contributing to the longitudinal findings. (d) Are the changes found longitudinally also found when men of the same age are compared at different times of measurement? This comparison provides information as to whether longitudinal changes are due to maturation or to changes taking place in the culture over the period of measurement.

---

[1] The contributions of the following persons in administering and scoring the GZTS are gratefully acknowledged: Sally Stram, Patricia Lamb, Robert Walker, Anne Young, Jack Zaner, Richard Mathias, Don Reynolds, Gail Schroeder, Barbara Donovan, Alice Musk, Patricia Allen, Marcia Schwartz, Darrell Gray, Joan King, Steven Kanis, Jean Onufry, and Marian Hedrick.

[2] Requests for reprints should be sent to David Arenberg, Gerontology Research Center, NIA, Baltimore City Hospitals, Baltimore, 21224.

[3] This study was initiated in 1958 by Dr. Melvin Davidoff who selected the GZTS and supervised the data collection until 1960.

These analyses represent attempts to describe changes in adult personality which are due to maturation, rather than cultural influences. Although the results from each type of analysis have implications for the interpretation of results from the other analyses, results will be reported separately for each analysis.

CROSS-SECTIONAL ANALYSES

METHOD

Subjects. — Participants in the Baltimore Longitudinal Study are a highly select group of male volunteers, predominantly white, who agree to return for testing at fixed intervals. The vast majority have earned at least a college degree, work in (or are retired from) scientific, professional, or managerial positions, and are in good health (Stone & Norris, 1966). Data in this paper are from responses of 915 participants who entered the Study from late 1958 to mid-1974. For several analyses the sample was dichotomized into those given a GZTS before and after mid-1968.

Procedure. — Each subject was given the standard GZTS instructions individually and completed the questionnaire during the remainder of his 3-day visit to the Gerontology Research Center. For each item, subjects choose "Yes", "No", or "?". Each scale consists of 30 items, but only "Yes" and "No" responses contribute to the scale score. A measure was invalidated for any scale with more than three "?" responses, a procedure suggested by Guilford and Zimmerman (1949).

The total sample was divided into two subsamples according to the date when the first GZTS was administered. The first group (Sample A) was comprised of 605 men who completed their first GZTS between October, 1958 and June, 1968. The age range for Sample A was 17 to 98 years old. The replication group (Sample B) included 310 men whose GZTS was administered from July, 1968 through June, 1974, and their age range was 18 to 90 years. In both Sample A and Sample B, the N for each scale was reduced by the number of measures that were invalidated due to excessive use of "?" responses for that scale. A product-moment correlation between age and each scale score was obtained.

Traits. — General Activity (G) predominantly involves the pace of activity. A high score characterizes an "energetic, rapid-moving, rapid-working person who likes action and may sometimes be impulsive." (Guilford & Zimmerman, 1956).

Restraint (R) predominantly involves a dimension of emotional maturity, with serious-minded at one pole and carefree at the other. A high score characterizes a self-controlled, persevering, responsible person who thinks before taking action, is not happy-go-lucky, and does not crave excitement.

Ascendance (A) is, in part, a dimension of social assertiveness. A high score characterizes an outspoken person who is prepared to take charge, stands up for his rights, and does not avoid verbal confrontation.

Sociability (S) is predominantly gregariousness or a dimension of seeking versus avoiding social interaction. A high score characterizes an outgoing person who is comfortable with others, prefers not to be or work alone, and is not shy.

Emotional Stability (E) is predominantly a dimension of uniform and positive outlook at one pole and changeable (mostly negative) mood at the other. A high score characterizes an optimistic person typically in good spirits who is not moody and rarely worries.

Objectivity (O) is a dimension involving a realistic outlook. A high score characterizes a person who believes he receives his due, who is not self-centered, and who does not feel unjustifiably blamed, criticized, or talked about.

Friendliness (F) is predominantly a dimension with agreeableness at one pole and hostility at the other. A high score characterizes a compliant person who is not contemptuous of others, does not resent being given orders, and is not easily aroused to belligerence or aggressive behavior.

Thoughtfulness (T) is a dimension of reflectiveness. A high score characterizes an introspective, meditative person who is given to analytic and evaluative thinking about behavior and ponders over the past.

Personal Relations (P) is a dimension of tolerance and cooperativeness. A high score characterizes a trustful person who thinks well of people and of institutions and who is not given to fault-finding or self-pity.

Masculinity (M) is a composite dimension predominantly involving "masculine" interests. A high score characterizes a person who is comfortable with guns and hunting and who has little compassion for animals, does not cry or express emotion easily, and is not given to ready feelings of fear or disgust.

Guilford et al., (1976) present a more detailed description of each scale, along with clinical interpretations.

RESULTS

Means and correlations with age for all scales appear in Table 1. Because Sample B correlations so closely replicated those for Sample A, the results reported below describe both samples. Five scales were related to age ($p < .01$) in both samples: General Activity, Ascendance, and Masculinity were negatively correlated with age, and Restraint and Friendliness were positively correlated with age. No statistically significant correlation ($p > .01$) was found for the other scales.

Table 1. Cross-sectional Data — Mean First Tests by Age Decade.

Age Decade =			20		30		40		50		60		70		80	
	r	(N)	N	$\overline{X}$	N	$\overline{X}$	N	$\overline{X}$	N	$\overline{X}$	N	$\overline{X}$	N	$\overline{X}$	N	$\overline{X}$
Total N																
Sample A			22		88		154		143		106		78		10	
Sample B			50		66		44		45		40		51		12	
Scale																
(G) Sample A	-.17**	(552)	20	18.05	77	18.78	141	17.53	134	17.31	98	16.92	72	14.94	8	12.75
Sample B	-.25**	(290)	45	19.13	65	18.62	42	18.26	40	18.07	36	17.17	48	14.65	12	14.17
(R) Sample A	.28**	(560)	20	17.40	78	18.01	142	19.61	135	20.42	102	20.75	71	21.93	8	20.75
Sample B	.26**	(285)	44	18.11	63	17.41	42	18.81	39	19.62	37	20.35	47	19.89	11	22.45
(A) Sample A	-.19**	(549)	20	18.35	81	17.44	138	16.94	136	16.22	96	15.71	66	14.35	8	13.25
Sample B	-.22**	(282)	46	18.11	65	18.26	42	17.45	39	17.82	32	15.03	44	15.50	12	14.50
(S) Sample A	-.09*	(538)	21	20.71	77	21.26	137	19.50	132	18.84	94	19.50	67	18.58	8	19.87
Sample B	-.07	(280)	46	19.24	64	20.72	40	18.57	40	19.22	32	18.34	46	19.50	10	17.50
(E) Sample A	.05	(576)	21	21.29	81	20.25	148	20.52	141	20.72	102	21.15	72	21.28	7	18.57
Sample B	.00	(287)	46	21.09	64	21.25	42	21.10	39	20.36	33	19.42	49	21.67	12	20.83
(O) Sample A	.01	(552)	21	21.10	80	20.97	140	20.45	133	20.62	100	21.04	70	21.34	6	16.33
Sample B	-.01	(278)	47	20.51	62	20.29	43	20.19	39	20.23	31	18.45	44	20.57	10	21.30
(F) Sample A	.17**	(556)	21	14.95	81	16.42	137	17.88	135	18.01	101	18.29	71	19.13	7	17.71
Sample B	.18**	(278)	45	15.33	62	15.82	41	16.05	39	17.00	34	16.76	44	18.05	11	18.55
(T) Sample A	.09*	(561)	21	19.24	80	18.60	136	18.72	136	19.51	104	19.64	73	19.79	8	18.62
Sample B	.11	(286)	46	18.65	62	17.90	42	17.79	39	18.64	36	18.61	47	19.57	12	21.42
(P) Sample A	.06	(539)	21	19.38	77	21.25	136	22.35	134	22.74	95	21.42	68	22.32	6	20.83
Sample B	.14*	(257)	46	18.33	57	18.91	39	19.85	37	19.59	29	19.62	41	20.00	6	21.17
(M) Sample A	-.17**	(563)	20	20.65	81	20.80	145	20.36	138	20.11	98	19.17	71	19.15	8	18.62
Sample B	-.25**	(291)	48	20.83	62	20.45	43	19.93	39	20.26	36	18.89	49	18.33	12	17.33

Note: Two teenagers and two men in their 90s do not appear in decade means for Sample A; one teenager and one 90-year-old do not appear in decade means for Sample B.

*$p < .05$.

**$p < .01$.

Decade means for General Activity decreased consistently after the 30s, and a substantial decrease was found after age 70 in both samples. Means for Ascendance decreased monotonically in Sample A, but less consistently in Sample B. The decrease in Masculinity was not apparent in either sample until age 60. Decade means for Restraint and Friendliness increased monotonically through the 70s in Sample A, but somewhat less consistently in Sample B.

Conventional Longitudinal Analyses

Method

Sample attrition. — Table 2 contains an accounting of subject loss. The 336 men included in longitudinal analyses ranged from 20 to 81 years of age when tested initially; they returned for their second GZTS 5.6 to 9.9 years later. In order to assess the effects of subject loss, the Sample A men were divided into two groups: repeats and nonrepeats. Repeats included the 336 men who appeared in both Sample A and the longitudinal sample; nonrepeats included 269 men who only appeared in Sample A. Because the number of subjects in several age decades was small (less than 10), the men were divided into three age groups: (a) young (17 - 39 years), (b) middle (40 - 59 years), and (c) old (60 - 98 years). Data were analyzed using a 2 × 3 unweighted-means analysis of variance (see Winer, 1971).

Table 2. Sources of Subject Loss for All Scales Combined.[a]

	Age Decade at First Test									All
	10	20	30	40	50	60	70	80	90	Ages
Sample A N =	2	22	88	154	143	106	78	10	2	605
Reasons:										
Died	0	0	1	4	7	9	20	4	2	47
Quit Program	1	8	19	30	27	21	19	5	0	130
Excluded[b]	1	2	18	22	10	13	5	0	0	71
Other[c]	0	2	3	9	2	2	3	0	0	21
Total Lost =	2	12	41	65	46	45	47	9	2	269

[a] The N for each scale was further reduced by the number of men whose first or second test was invalid due to excessive use of question marks.

[b] Second GZTS obtained after July 1, 1974, repeat interval exceeded 9.9 years, or additional exposure to GZTS between first and longitudinal measures.

[c] Includes men who were still active in program, but had not visited the Center during time period in which they were eligible for a second GZTS, and also men who refused to fill out the GZTS.

Longitudinal change. — To determine whether scores for a scale changed, the mean of the changes for all subjects was compared with zero. Furthermore, to determine whether magnitude of change was related to age, a part correlation between age and the residual of the second measure adjusted for the first measure was computed (see DuBois, 1957).

Results

Sample attrition. — Results of the analyses of variance to determine the effects of sample attrition appear in Table 3. The difference between samples (repeats vs. nonrepeats) was statistically significant for Objectivity and Friendliness ($p < .01$), and for Ascendance, Emotional Stability, and Personal Relations ($p < .05$). Furthermore, the interaction between age group and sample was significant for General Activity ($p < .01$), and for Personal Relations ($p < .05$). Repeats in all age groups were lower in Ascendance and higher in Emotional Stability, Objectivity, and Friendliness than nonrepeats. In General Activity, the repeats were lower than the nonrepeats for the young groups, the two middle groups were about the same, and the repeats were higher than the nonrepeats for the old groups. Repeats in the young group were higher in Personal Relations than nonrepeats, but no difference was found in the middle and old groups.

It should be noted that the mean age of young nonrepeats was lower than that of young repeats (32.95 and 34.08 respectively), the middle nonrepeats and repeats had about the same mean age (49.26 and 49.74 respectively), and the mean age of old nonrepeats was higher than that of old repeats (71.28 and 67.56 respectively). The possibility that the scale score differences between repeats and nonrepeats were influenced by age differences between the two groups, was explored by comparing scale means for subgroups of repeats and nonrepeats which were approximately the same mean age. Although no statistical tests were performed, patterns similar to those discussed were found for all scales.

Table 3. Mean First Tests for Repeats vs. Nonrepeats.

Scale		N	Young	N	Middle	N	Old	F Ratios	
								Sample[a]	Age × Sample[b]
(G)	Nonrepeats	50	19.70	103	17.48	100	15.09	.08	4.66**
	Repeats	49	17.37	172	17.40	78	17.01		
(R)	Nonrepeats	50	18.10	107	20.00	101	21.31	.24	.08
	Repeats	50	17.74	170	20.01	82	21.12		
(A)	Nonrepeats	56	18.41	108	17.07	93	15.48	5.37*	.44
	Repeats	47	16.57	166	16.27	79	14.61		
(S)	Nonrepeats	52	21.17	107	19.50	95	18.91	.05	.40
	Repeats	48	20.79	162	18.96	74	19.47		
(E)	Nonrepeats	54	20.02	113	19.72	101	20.49	5.40*	.21
	Repeats	50	20.74	176	21.20	82	21.81		
(O)	Nonrepeats	57	20.42	110	20.02	95	20.22	8.50**	.33
	Repeats	46	21.57	163	20.88	81	21.91		
(F)	Nonrepeats	57	14.86	106	16.93	97	18.44	11.32**	2.04
	Repeats	47	17.70	166	18.60	83	18.87		
(T)	Nonrepeats	53	18.83	104	20.23	103	19.81	2.76	1.66
	Repeats	50	18.78	168	18.42	83	19.53		
(P)	Nonrepeats	52	19.56	106	22.40	101	21.66	5.36*	3.12*
	Repeats	48	22.15	164	22.63	68	21.91		
(M)	Nonrepeats	52	20.69	115	19.84	95	18.77	2.64	.46
	Repeats	51	20.78	168	20.51	82	19.57		

[a] $df = 1$

[b] $df = 2$

*$p < .05$.

**$p < .01$.

Longitudinal change. — The longitudinal analyses of change showed an over-all decline on General Activity (t (298) = $-3.50, p < .01$). Furthermore, the magnitude of change was related to age. The part correlation between age and change score for General Activity was $-.23$ (F (1,296) = $16.81, p < .01$). Inspection of Table 4 shows that 20-year-olds increased, whereas men over 50 declined, in General Activity.

Four other scales showed overall declines from first to second test: Friendliness (t (295) = -2.35, $p < .05$), Thoughtfulness (t (300) = $-2.33, p < .05$), Personal Relations (t (279) = -3.82, $p < .01$), and Masculinity (t (300) = -2.54, $p < .01$). However, the magnitude of change was not systematically related to age on any of these scales. All age groups declined in Thoughtfulness and Personal Relations, all groups except the 30s declined in Friendliness, and all age groups except the 30s and 60s declined in Masculinity.

In summary, declines were found from first test to second test for General Activity, Friendliness, Thoughtfulness, Personal Relations, and Masculinity, but the magnitude of change varied linearly with age only on General Activity. Differences between repeats and nonrepeats were found in General Activity, Ascendance, Emotional Stability, Objectivity, Friendliness, and Personal Relations.

SEQUENTIAL ANALYSES OF INDEPENDENT SAMPLES

METHOD

Although data collection was not planned for sequential analyses, an attempt was made to analyze subsets of the cross-sectional data cross-sequentially and time-sequentially. The first time-of-measurement sample (Group 1) in both designs consisted of the 518 men from Sample A born between the years 1892 and 1939. The age range for Group 1 was 24 to 76. For the cross-sequential analyses, the second time-of-measurement sample (Group 2-CS)

Table 4. Longitudinal Data — Mean First and Second Tests by Age Decade.

Age at First Test =	N	20	N	30	N	40	N	50	N	60	N	70	N^a	All Ages
Scale														
(G) First Test	9	16.44	40	17.58	83	17.60	89	17.20	52	17.21	25	16.36	299	17.29
Second Test		18.11		17.48		17.59		16.12		15.17		15.08		16.54
Change		1.67		-.10		-.01		-1.08		-2.04		-1.28		-.76**
(R) First Test	9	16.78	41	17.95	82	19.49	88	20.49	57	20.74	24	22.21	302	19.93
Second Test		16.78		18.39		18.92		20.53		20.70		21.33		19.79
Change		.00		.44		-.57		.05		-.04		-.88		-.14
(A) First Test	7	17.43	40	16.43	80	16.85	86	15.73	55	15.09	23	13.13	292	15.87
Second Test		17.57		16.53		17.20		15.49		14.89		13.17		15.87
Change		.14		.10		.35		-.24		-.20		.04		.00
(S) First Test	9	19.33	39	21.13	80	19.83	82	18.12	49	19.29	24	19.54	284	19.41
Second Test		20.44		20.77		20.04		17.92		18.18		19.50		19.19
Change		1.11		-.36		.21		-.21		-1.10		-.04		-.21
(E) First Test	9	22.78	41	20.29	85	21.42	91	20.99	54	21.19	27	23.00	308	21.29
Second Test		23.56		20.76		21.39		21.59		20.56		22.67		21.39
Change		.78		.46		-.04		.60		-.63		-.33		.10
(O) First Test	8	22.38	38	21.40	81	21.15	82	20.62	55	21.42	26	22.96	290	21.28
Second Test		21.50		21.13		21.20		21.01		21.53		22.04		21.28
Change		-.88		-.26		.05		.39		.11		-.92		.00
(F) First Test	8	18.00	39	17.64	81	18.70	85	18.49	57	18.12	25	20.72	296	18.53
Second Test		17.75		17.64		17.83		18.06		18.05		19.80		18.05
Change		-.25		.00		-.88		-.44		-.07		-.92		-.48*
(T) First Test	9	17.67	41	19.02	81	18.11	87	18.71	59	19.39	23	19.83	301	18.79
Second Test		15.33		18.24		17.57		18.64		18.92		19.17		18.31
Change		-2.33		-.78		-.54		-.07		-.48		-.65		-.48*
(P) First Test	9	20.22	39	22.59	79	22.77	85	22.51	46	21.26	21	23.62	280	22.38
Second Test		19.89		21.46		22.39		21.73		20.52		21.24		21.55
Change		-.33		-1.13		-.38		-.78		-.74		-2.38		-.82**
(M) First Test	9	20.11	42	20.93	81	20.44	87	20.56	57	19.72	24	19.38	301	20.30
Second Test		19.78		21.07		19.84		19.95		19.72		18.42		19.89
Change		-.33		.14		-.61		-.61		.00		-.96		-.41**

[a] All ages includes one man in his 80s who does not appear in decade means.

*$p < .05$.

**$p < .01$.

consisted of 238 men from Sample B born during the same time period, but ranging in age from 30 to 81. Group 2-TS in the time-sequential analyses included the 246 Sample B men who were born between 1900 and 1947. They were similar in age to Group 1 (from 24 to 72 years). In each design the mean time of measurement for the late sample was approximately 7 years later than the early sample. The *N* for each scale was reduced by the number of men whose measures were invalid for that scale.

In both designs each group was divided into six consecutive birth cohorts, with each cohort including men born during an 8-year period. In the cross-sequential design, men born during the same time period (cohort) were compared at different ages. Assuming no sample biases (e.g., selective survival in the late sample), differences found between Group 1 and Group 2-CS are estimates of age changes, but also could be due to cultural change from 1958-1968 to 1968-1974. The time-sequential design provided a comparison of men who were the same age but born at different times. Under the assumption of no sample biases, differences between Group 1 and Group 2-TS in this design could be attributed to cohort differences, cultural change from the first period of measurement to the second, or both. A 2 × 6 unweighted-means analysis of variance was conducted for each design on the mean first tests for each scale.

Table 5. Means for Sequential Analyses of Independent Samples.

		Date of Birth														Unweighted Means		
		N	1940-1947	*N*	1932-1939	*N*	1924-1931	*N*	1916-1923	*N*	1908-1915	*N*	1900-1907	*N*	1892-1899	Group 1	Group 2-CS[c]	Group 2-TS[d]
Scale																		
(G)	Group 1		—	19	18.79	70	18.44	99	18.25	128	16.42	82	18.23	77	16.13	17.71	17.41	18.03
	Group 2	47	18.49	49	19.39	35	18.43	36	18.81	25	16.52	37	16.57	41	14.76			
(R)	Group 1		—	19	17.63	73	18.33	100	19.11	125	20.46	85	20.05	79	21.17	19.46	19.17	19.07
	Group 2	45	18.78	47	17.23	35	18.17	36	19.42	24	20.33	39	20.51	39	19.36			
(A)	Group 1		--	19	18.37	75	16.87	95	18.27	128	15.20	82	16.88	77	15.46	16.84	17.00	17.47
	Group 2	47	18.15	49	18.82	35	17.40	36	17.53	23	17.78	35	15.14	36	15.33			
(S)	Group 1		—	19	21.47	72	20.82	97	20.52	122	18.00	84	19.98	72	18.99	19.96	19.43	19.34
	Group 2	48	19.33	49	21.45	34	19.27	35	18.80	24	19.17	35	18.00	38	19.92			
(E)	Group 1		—	19	21.58	75	19.91	106	20.79	131	20.34	85	21.11	82	21.50	20.87	20.84	20.88
	Group 2	48	21.90	48	21.00	35	21.71	36	21.06	23	19.52	37	20.08	40	21.65			
(O)	Group 1		—	20	22.75	74	20.65	97	20.42	127	20.70	82	20.90	78	21.15	21.10	19.98*	20.14*
	Group 2	47	20.96	48	20.27	35	20.43	36	20.39	23	19.13	31	19.65	39	20.00			
(F)	Group 1		—	20	16.45	73	15.88	97	17.55	127	18.62	83	17.71	81	18.43	17.44	16.70	16.47*
	Group 2	46	15.85	48	15.83	33	16.30	36	16.39	23	16.96	36	17.50	37	17.24			
(T)	Group 1		—	20	18.85	73	18.86	94	18.57	129	18.77	84	20.27	82	19.92	19.21	18.43	18.23*
	Group 2	47	18.38	48	18.23	33	16.67	36	18.67	23	18.35	39	19.10	39	19.56			
(P)	Group 1		—	20	22.40	73	21.18	93	22.33	127	22.61	81	22.11	74	21.32	21.99	19.47**	19.42**
	Group 2	43	19.56	47	18.83	31	19.61	34	19.68	21	19.10	30	19.73	35	19.86			
(M)	Group 1		—	19	21.32	76	20.65	101	20.52	128	20.33	86	18.87	76	19.87	20.26	19.54*	20.07
	Group 2	49	21.35	48	20.15	33	19.94	38	20.40	24	19.83	37	18.76	41	18.15			
Mean Ages:																		
	Group 1		—		29		36		44		51		59		67			
	Group 2		29[a]		35		43		51		58[b]		67		74	48	55	47

[a] Scale S mean age was 28.

[b] Scale M mean age was 59.

[c] Group 2-CS (cross-sequential analyses) excludes Group 2 men born from 1940-1947.

[d] Group 2-TS (time-sequential analyses) excludes Group 2 men born from 1892-1899.

*$p < .05$.

**$p < .01$.

RESULTS

Means for both the cross-sequential and time-sequential designs appear in Table 5. The cross-sequential comparisons are within cohorts and can be seen by comparing column means (e.g., Groups 1 and 2 at 1932-1939 cohort). Comparisons for the time-sequential analyses are within ages and can be seen in diagonal comparisons (e.g., Group 1 at 1932-1939 cohort and Group 2 at 1940-1947 cohort).

A significant time effect was found (i.e., Group 1 was different from both 2-CS and 2-TS) for Objectivity in both the cross-sequential and time-sequential designs (F (1,678) = 6.62, $p < .05$, and F (1,686) = 5.20, $p < .05$ respectively), and also for Personal Relations (F (1,654) = 32.99, $p < .01$ in cross-sequential analysis, and F (1,662) = 34.70, $p < .01$ in time-sequential design). In the cross-sequential analyses, Group 2-CS men were lower in every cohort on Objectivity and Personal Relations than Group 1 men. The time-sequential analyses showed that except for the group whose mean age was in the 40s, all ages in Group 2-TS were lower in Objectivity than in Group 1, and on Personal Relations every age group in Group 2-TS was lower than in Group 1.

A time effect for Masculinity was seen only in the cross-sequential analysis (F (1,695) = 5.43, $p < .05$). Every cohort in Group 2-CS was lower in Masculinity than in Group 1.

Friendliness and Thoughtfulness showed time effects only in the time-sequential analyses (F (1,691) = 4.23, $p < .05$ and F (1,696) = 5.66, $p < .05$ respectively). Both Friendliness and Thoughtfulness were lower for each age group in Group 2-TS than in Group 1.

In summary, Groups 2-CS and 2-TS were lower than Group 1 men on Objectivity and Personal Relations. In the cross-sequential design Group 2-CS men were lower on Masculinity than Group 1 men. In the time-sequential design, Group 2-TS was lower on Friendliness and Thoughtfulness than Group 1. No significant interaction was found for any scale in either design.

DISCUSSION

Let us begin with the basic findings and conclusions of this study. Longitudinal changes were found for five scales, but only two scales were interpreted as showing maturational effects. Beginning at age 50, preference for rapidly-paced activity declined; and at all ages men declined in masculine interests. The tendency toward analytic and evaluative thinking declined due to cultural influences during the time period spanned by the study, as did tolerance, cooperativeness, and trust in other people and institutions. A long-term cultural drift resulted in an increased tendency toward belligerence and hostility. The findings for two other scales were interpreted as showing generational differences. Later-born men were less serious-minded, persevering, and responsible, and more socially assertive than early-born men.

Several methodological points should be kept in mind in considering the conclusions of this study. Because age is not an experimentally manipulated variable, age differences and even age changes may be due to other than maturational effects. As noted above, cross-sectional age differences may be due to maturational effects or to cohort differences. Similarly, longitudinal age changes and within-cohort age differences in cross-sequential analyses may be due to maturational effects or to sociocultural changes over the period of measurement. In synthesizing the results of all analyses into a meaningful interpretation, an attempt was made to attribute the results for each age-related scale to one effect: maturation, generation, or sociocultural change specific to the period of measurement. Due to the confound between age and environmental effects, the findings are specific to those generations studied and the particular time-of-measurement period; but the fact that a wide span of birth cohorts was studied over a 16-year period increases the likelihood that similar results would be found in other samples at other times.

A summary of the findings for the six scales which showed interpretable effects appears in Table 6.

Maturation — Maturational change alone could account for the findings for two scales: General Activity and Masculinity. The decline in Masculinity shown in repeated-measures analyses was confirmed in cross-sequential analyses, and Masculinity also decreased with

Table 6. Summary of Results From All Analyses.

	Cross-sectional	Conventional Longitudinal	Cross-sequential	Time-sequential
Maturation				
General Activity	Decreased	Declined	Decreased (N.S.[a])	—
Masculinity	Decreased	Declined	Decreased	—
Cultural Change				
Thoughtfulness	—	Declined	Decreased (N.S.)	Decreased
Personal Relations	—	Declined	Decreased	Decreased
Generation				
Restraint	Increased	—	—	Decreased (N.S.)
Ascendance	Decreased	—	—	Increased (N.S.)
Long-term Cultural Change				
Friendliness	Increased	Declined	Decreased (N.S.)	Decreased

[a] N.S. = not statistically significant

age cross-sectionally. No decline in Masculinity was found when men of the same age, but tested at different times, were compared (time-sequential analyses). Whereas Masculinity declined consistently across the age span. General Activity showed change which was related to initial age. Conventional longitudinal results showed that men in their 20s increased on General Activity, whereas men over 50 years of age declined. Cross-sequential decreases, although smaller, were consistent with these repeated-measures findings. Perhaps the magnitude of the longitudinal (within-subject) decline was affected by the relatively high scores of the old men who were included in the repeated-measures sample. Both the time-sequential results and the cross-sectional findings were consistent with maturational decline in General Activity.

As men grew older, then, they conformed less to the classical pattern of male interests. Furthermore, after the age of 50, they became less disposed to perform at a rapid pace or to be physically active.

Cultural change. — Thoughtfulness and Personal Relations showed declines in both the conventional longitudinal and cross-sequential analyses. The finding of declines in time-sequential analyses and the absence of cross-sectional differences suggested cultural change during the measurement period as an explanation. Therefore, men became less interested in reflective and introspective thought, and also became less tolerant and cooperative during the time frame of the current study.

Generation. — Restraint and Ascendance showed age differences in both cross-sectional analyses, but neither measure changed with age longitudinally. Later-born cohorts were lower in Restraint and higher in Ascendance than early-born cohorts. The time-of-measurement variable in the time-sequential analyses, although not statistically significant, was consistent with the cohort interpretation for these two scales. These data indicate that successive generations were somewhat less responsible and self-controlled, and were more self-assertive.

Generation and cultural change. — The results for Friendliness were complex and cannot be explained by a single effect. This measure increased with age cross-sectionally in both samples, but declined longitudinally within subjects. Furthermore, the cross-sequential comparisons within cohorts, although not statistically significant, were consistent with the changes (declines) in the conventional longitudinal analysis. These results indicated that Friendliness declined over the period of measurement in this study. However, the time-sequential differences between subjects of the same age suggest that this change was not maturational but rather a cultural change, possibly specific to these periods of measurement. In addition, the cross-sectional results require a cohort interpretation; that is, later-born cohorts were lower in Friendliness than early-born cohorts. Therefore, a cohort effect from the past and a cultural change during the period of measurement would account for the complex results for this scale. Another way these results could occur is that individuals decline in Friendliness as they grow older (rather than changing only during the period covered in this study), and successive cohorts begin at lower and lower levels. Although the decline could be maturational, the more parsimonious interpretation is that a long-term cultural drift was responsible for generation differences as well as change within individuals. It is concluded, therefore, that successive cohorts and individuals became more easily aroused to hostility and tended to be less agreeable.

No effect. — Analyses did not show age differences or age changes for Sociability or Emotional Stability. Objectivity decreased in both sequential designs; however, no change was found in the repeated-measures analyses. Therefore, gregariousness, consistency of positive outlook, and tendency to take a realistic outlook were not affected by maturation, generation, or cultural change.

The conclusions of this study can be related to several prominent thoughts in personality and aging. The Disengagement Theory formulated by Cumming and Henry (1961) has received much attention by gerontologists. Although part of this theory has been challenged by subsequent evidence, the hypothesis of decreased social interaction in later life has received wide support. The GZTS scale which comes closest to measuring this dimension is Sociability. The finding of no maturational decline in Sociability runs counter to this well-accepted hypothesis.

Another hypothesis which is closely related to disengagement is that people become more introverted in later life (Botwinick, 1973; Neugarten, 1977). If Thoughtfulness can be considered a measure of introversion, then current findings do not support this hypothesis. Thoughtfulness was found to decline due to cultural influences rather than maturation.

Gutmann (1977) has proposed that as men age they move from active toward passive mastery of the environment. Active mastery is a broad concept which encompasses dominance, instrumentality, and participation in traditionally male activities. Gutmann's hypothesis was partially supported by findings in this study in that Masculinity showed maturational declines. However, Ascendance showed generational, but not maturational effects. Although men declined in masculine interests, they did not decline in social assertiveness.

Both Botwinick (1973) and Neugarten (1977) discuss the importance of health and economic factors as contributors to personality change with age. The findings of the current study can be interpreted as further support. Participants in the Baltimore Longitudinal Study as a group are highly educated, in good health, and economically advantaged. Perhaps when health and economic resources are maintained the "turning inward" process (Botwinick, 1973) may be forestalled.

The analyses which have been reported in this study were concerned with quantitative change on the GZTS scales. Although several studies have reported cross-sectional data concerning age and personality structure (Bendig, 1960; Costa & McCrae, 1976), the issue of age changes in personality structure remains to be investigated.

An important consideration in analysis of the GZTS is loss of data due to use of the question-mark response. The samples in all analyses were biased to the extent that noncommital responding is related to scale scores on the GZTS. Although beyond the scope of the current analyses, failure to complete a questionnaire or to provide a scoreable response are interesting and potentially important behaviors which have received little attention in the personality and aging literature.

SUMMARY

Cross-sectional and conventional longitudinal comparisons for each of the ten scales of the GZTS were made for the men in the Baltimore Longitudinal Study. Early and late subsamples of the cross-sectional sample were regrouped into both cross-sequential and time-sequential designs to provide additional information concerning age differences and age changes in these measures. Participants were volunteers who ranged in age from 17 to 98 years and were, for the most part, educated, in relatively good health, and employed in (or retired from) professional or managerial positions. Care should be taken in generalizing findings from such a select group to populations which differ in these characteristics.

Cross-sectional increases with age were found for Restraint and Friendliness, and decreases were found for General Activity, Ascendance, and Masculinity.

Longitudinal changes were found in five scales. General Activity showed an over-all decline, and the magnitude of change varied according to age group. The men in their 20s actually increased somewhat, and only the means for the men over 50 declined. The measures of Friendliness, Thoughtfulness, Personal Relations, and Masculinity also declined for those men who had two valid measures approximately 7 years apart.

The five scales which showed changes in the conventional longitudinal analyses seemed to fall into three types. The declines in General Activity and in Masculinity appeared to be maturational changes; the declines in Thoughtfulness and Personal Relations appeared to be cultural changes specific to the period spanned by the two times of measurement; and the declines in Friendliness appeared to be attributable to a long-term cultural drift.

The age differences found cross-sectionally for Restraint and Ascendance were consistent with generation effects rather than maturation effects. Successive birth cohorts were lower in Restraint and higher in Ascendance, but changes were not found within individuals and differences were not found within cohorts.

REFERENCES

Baltes, P. B. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, 1968, *11*, 145-171.

Bendig, A. W. Age differences in the interscale factor structure of the Guilford-Zimmerman Temperament Survey. *Journal of Consulting Psychology*, 1960, *24*, 134-138.

Botwinick, J. *Aging and behavior*. Springer, New York, 1973.

Costa, P. T. Jr., & McCrae, R. R. Age differences in personality structure: A cluster analytic approach. *Journal of Gerontology*, 1976, *31*, 564-570.

Cumming, E., & Henry, W. E. *Growing old*. Basic Books, New York, 1961.

DuBois, P. H. *Multivariate correlational analysis*. Harper & Brothers, New York, 1957.

Guilford, J. P., & Zimmerman, W. S. *The Guilford-Zimmerman Temperament Survey: Manual of instructions and interpretations*. Sheridan Supply Company, Beverly Hills, 1949.

Guilford, J. P., & Zimmerman, W. S. Fourteen dimensions of temperament. *Psychological Monographs*, 1956, *70* (10, Whole No. 417).

Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. *The Guilford-Zimmerman Temperament Survey handbook*. EdITS Publishers, San Diego, 1976.

Gutmann, D. The cross-cultural perspective: Notes toward a comparative psychology of aging. In J. E. Birren & K. W. Schaie (Eds.) *Handbook of the psychology of aging*. Van Nostrand Reinhold, New York, 1977.

Kelly, E. L. Consistency of the adult personality. *American Psychologist*, 1955, *10*, 659-681.

Neugarten, B. L. Personality and aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. Van Nostrand Reinhold, New York, 1977.

Schaie, K. W. A general model for the study of developmental problems. *Psychological Bulletin*, 1965, *64*, 92-107.

Schaie, K. W. A reinterpretation of age related changes in cognitive structure and functioning. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology*. Academic Press, New York, 1970.

Schaie, K. W., & Parham, I. A. Stability of adult personality traits: Fact or fable? *Journal of Personality and Social Psychology*, 1976, *34*, 146-158.

Stone, J. L., & Norris, A. H. Activities and attitudes of participants in the Baltimore Longitudinal Study. *Journal of Gerontology*, 1966, *21*, 575-580.

Titus, H. E., & Goss, R. G. Psychometric comparison of old and young supervisors. *Psychological Reports*, 1969, *24*, 727-733.

Wagner, E. E. Differences between old and young executives on objective psychological test variables. *Journal of Gerontology*, 1960, *15*, 296-299.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). McGraw-Hill, New York, 1971.

Woodruff, D. S., & Birren, J. E. Age changes and cohort differences in personality. *Developmental Psychology*, 1972, *6*, 252-259.

Journal of Gerontology
1983, Vol. 38, No. 2, 162-180

A Longitudinal Study of Nutritional Intake in Men[1,2]

Virginia K. Elahi, MPH,[4] Dariush Elahi, PhD,[4]
Reubin Andres, MD,[4] Jordan D. Tobin, MD,[4]
Marlene G. Butler, MS, RD,[4] and Arthur H. Norris, MS[3,4]

Seven-day dietary diaries were provided by 180 male participants in the Baltimore Longitudinal Study of Aging during each of three time periods (1961 to 1965, 1966 to 1970, and 1971 to 1975). These men are a highly educated, upper-middle class group. At the time of their first diary, they were aged 35 to 74 years. The data were analyzed for aging, cohort, and time effects on diet by utilizing three types of research designs concurrently: cross-sectional, longitudinal, and time series. The nutrients considered were calories, protein, carbohydrate, fat, saturated fatty acids, polyunsaturated fatty acids, and cholesterol. Aging had a negative effect on intake of calories, fat, saturated fatty acids, and cholesterol. Cohort effects were not observed for any of these nutrients. Over time, intake of carbohydrates and cholesterol declined, while intake of polyunsaturated fatty acids rose.

Key Words: Cohort effects, Aging, Time effects, Calories, Protein, Carbohydrate, Fat, Cholesterol, Saturated fatty acids, Polyunsaturated fatty acids, Diet

THE importance of diet in the maintenance of good health is widely acknowledged. Schlenker et al. (1973) have reviewed a number of studies that relate dietary habits to health and longevity and have suggested that the rate of aging may be affected by long-term food habits. There are very little data, however, on the levels of nutrient intakes in humans and on the changes in intakes with increasing age and over time.

Cross-sectional data on diet, which show how nutrient intakes vary with age among people observed at a given point of time, are available from several sources, including the Framingham Study for the late 1950s (Kannel & Gordon, 1970), the Ten-State Nutrition Survey in 1968 to 1970 (U.S. Department of Health, Education, and Welfare, DHEW, 1972), and the Health and Nutrition Examination Survey (HANES) in 1971 to 1974 (U.S. DHEW, 1977). However, the Framingham and DHEW data are limited in age range and/or nutrients that they cover. Cross-sectional data from the Baltimore Longitudinal Study of Aging (BLSA) in the early 1960s have been presented by McGandy et al. (1966).

Data on changes in the diet of Americans over the past 70 years can be obtained from the U.S. Department of Agriculture (USDA), which has made annual estimates of the various types of food available for consumption in the market (see Brewster & Jacobson, 1978, for a summary of the USDA reports). However, these data represent average-per-person consumption for the entire United States population. Furthermore, they do not consider changes in the characteristics of the population (e.g., age distribution, ethnic composition) that have undoubtedly affected patterns of consumption.

Longitudinal data on dietary intake are available from the studies by Steinkamp et al. (1965) in California and Garcia et al. (1975) in Iowa. Both studies began in 1948 and ended in the 1960s. But the sample in Steinkamp et al. was limited to individuals who were at least 50 years old when the study began and Garcia et al. obtained longitudinal data on only 28 women.

The aim of the present study is to provide descriptive data on nutrient intake among a group of

---

[1] This paper was presented in part at the 31st Annual Scientific Meeting of the Gerontological Society in Dallas, Texas, November, 1978.

[2] The nutrient intake study was initiated by Drs. Nathan W. Shock, Arthur E. Rikli, Robert B. McGandy, and Charles H. Barrows, Jr. Collection of the dietary data in the early part of the study was supervised by Marjorie C. Zukel, Alexandria Spanias, and Alla Meredith. We express our gratitude to them for their foresight in undertaking this difficult project and for the meticulous attention paid to data collection and coding. We are indebted to the participants of the Baltimore Longitudinal Study of Aging for their splendid cooperation in providing detailed data. The program used to convert food intakes to nutrient intakes was adapted by Dr. Ruth Brennan of the St. Louis County, MO, Health Department and the conversions were made under her supervision at Washington University. We also thank Helen Burns and Dolores Wehner for excellent secretarial assistance, and Paul Ciesla and Charlotte Adler for preparing the illustrations.

[3] Deceased December 1980.

[4] Clinical Physiology Branch, Gerontology Research Center, National Institute on Aging, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, at Baltimore City Hospitals, Baltimore, MD 21224. Address reprint requests to the first author.

middle-class, predominantly white, adult males over a 15-year period. The data consist of 7-day dietary diaries that were provided by 180 male participants in the BLSA between 1961 and 1975. The men ranged in age from 35 to 74 when they completed their first diary. This paper will describe how the diet of these men has changed as they have aged since the early 1960s. The nutrients that will be discussed are calories, protein, carbohydrates, fats, saturated fatty acids (SFA), polyunsaturated fatty acids (PFA), and cholesterol.

This study also attempts to identify aging, time, and cohort effects on nutrient intakes. In the least complicated terms, aging, time, and cohort effects can be defined as follows: An *aging* effect is present if the dependent variable is a function of age regardless of the subject's birthyear or period of observation. A *time* effect is present if the variable changes over time and the change is observed in all age groups. A *cohort* effect is present if the value of the dependent variable differs between birth cohorts regardless of age and period of observation.

Researchers interested in the effects of aging have usually employed one of two types of study designs. The first is a cross-sectional approach, in which a sample from a population is measured at one point in time. Cross-sectional differences between age groups, however, may represent differences between birth cohorts rather than aging effects. The second type of design is a longitudinal study which consists of observing one or more cohorts at several points in time. Differences seen between observation points are then attributed to the fact that the cohorts have aged. Longitudinal differences, however, may be due to temporal changes rather than to aging. Thus, it is difficult to establish the presence or absence of these effects when either of these approaches is the only one used.

The limitations of the conventional simple cross-sectional and longitudinal methods have been recognized by many researchers. Schaie (1965) proposed a sequential model based on analysis of variance; it was designed to separate effects due to chronological age, time of measurement, and year of birth. Baltes (1968) argued, however, that one cannot assign a unique effect to each of these independent variables as they are inherently confounded with each other (i.e., if the value of any two of the variables is given, the value of the third is fixed). He therefore suggested using a bifactorial model in which one of the three independent variables is ignored. More recently, Adam (1978) evaluated Schaie's work using illustrative data and concluded that the method cannot separate these three effects. She points out difficulties in the decision rules and ambiguities in the interpretation of results. Mason et al. (1973) proposed using multiple regression techniques to measure these three effects. By assuming that at least two age groups, two cohorts, or two time periods have identical effects, the authors believe that the problem of the interrelationship of the three independent variables is overcome. This method, however, has not been used extensively.

The present study attempts to differentiate aging, time, and cohort effects by utilizing three designs concurrently, each of which controls for a different effect: cross-sectional, longitudinal, and a third type designated as time series. We do not propose that this methodology provides a definitive solution to the problem of separating aging, time, and cohort effects. The mechanics of the analysis are relatively simple, however, and the method is very useful when there are pure aging, time, or cohort effects. When there are multiple effects it is impossible to unconfound the effects by this or any other method of which the authors are aware, and one can only estimate the probable nature of the effects.

RESEARCH METHODOLOGY

The hypothesis under consideration is that the level of intake of any specified nutrient is a function of age, cohort, and/or time. Intake is designated the dependent variable, and age, cohort, and time the independent or causal variables. The term independent should not be construed to mean that the causal variables in this analysis are unrelated to each other; as already stated, the value of any one of these variables is defined by the values of the remaining two.

The basis of the methodology in the present study is the age-time matrix shown in Figure 1. Age is on the vertical axis and is divided into intervals of width i. Time of observation is on the horizontal axis and is also divided into intervals of width i. The *width* of the age and time intervals *must* be equal in this methodology, but the number of age intervals does not have to equal the number of time intervals. (The number of age and time intervals shown here equal the numbers used in the nutritional intake study to be presented later.) Both the age and time ranges, however, should be sufficiently long to permit aging, time, and cohort effects to manifest themselves. One starts with a defined group of individuals at time t who range in age from a to $a + 7i$ and subsequently observes each person whenever i time units have passed. Theoretically, i may be equal to any quantity the researcher desires but, in practice, the magnitude of i is determined by the time necessary for changes to appear in the subjects. In studies of human aging, it is customary for i to equal 5 or 10 years. Since the age and time intervals are equal, all persons who are at a specified age a at time t will be at age $a + i$ at time $t + i$, and at age $a + 2i$ at time $t + 2i$. Thus, initial age groups or birth cohorts will move diagonally within this matrix. The numbers in the cells, C_k, are the mean values obtained for the dependent variable for cohort k at the specified age and time.

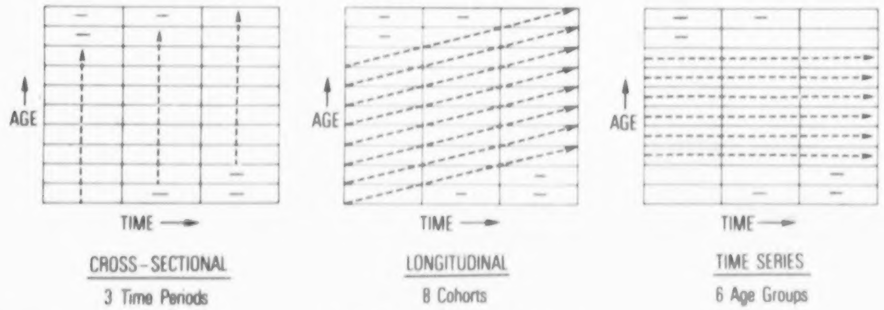
The three analytical perspectives employed in this method are illustrated in Figure 2 with the same age-time matrix. If one moves vertically in the matrix within a given time period, a *cross-sectional* approach is being used. On any given vertical line the effects of secular influences, that is, effects associated with the period of observation, are held constant as all individuals are being observed at the same time. But age is increasing, and the participants are from many different birth cohorts. If one moves diagonally to the right within this matrix, a *longitudinal* approach is being used. On any given diagonal the cohort is constant. But the age of the members of the cohort is increasing, and they are exposed to different period effects as time passes. In the third perspective, the *time series* approach, one moves horizontally to the right within a given age group. Age is constant on any given horizontal line. But the participants are from different birth cohorts and time is increasing.

Table 1 summarizes the differences in these three analytical perspectives. It can be seen that each of the independent variables — time, cohort, and age — is held constant by a different perspective. Because of this, if there is only a single or pure effect one can distinguish whether it is an aging, time, or cohort effect by considering the data from all three perspectives. Table 2 lists the types of perspectives that will show significant differences when a specified effect is present. If only a pure aging effect is operating, then we should find both the cross-sectional and longitudinal slopes significant; the time series analysis should show no significant slope. This follows because, first, age is the only independent variable that varies in both the cross-sectional and longitudinal perspectives, and, secondly, age is held constant in the time series approach. Similarly, if only a cohort effect is operating, then only the cross-sectional and time-series approaches should show significant differences; cohort is the one independent variable that varies in both these approaches, while the longitudinal perspective holds cohort effects constant. Likewise, if only a time effect is operating, then only the longitudinal and time series approaches should show

AGE			
$a + 9i$	—	—	C_8
$a + 8i$	—	C_8	C_7
$a + 7i$	C_8	C_7	C_6
$a + 6i$	C_7	C_6	C_5
$a + 5i$	C_6	C_5	C_4
$a + 4i$	C_5	C_4	C_3
$a + 3i$	C_4	C_3	C_2
$a + 2i$	C_3	C_2	C_1
$a + i$	C_2	C_1	—
a	C_1	—	—
	t	$t+i$	$t+2i$

TIME →

Figure 1. Age-time matrix, where t is time at first observation, a is youngest age observed at time t, i is width of age and time intervals, and c_k is value for birth cohort k at specified age and time.



CROSS-SECTIONAL
3 Time Periods

LONGITUDINAL
8 Cohorts

TIME SERIES
6 Age Groups

Figure 2. Analytical perspectives.

Table 1. Direction of Change in Independent Variables in Each Analytical Perspective

Analytical perspective	Independent variable		
	Time	Cohort	Age
Cross-sectional	CONSTANT	Increasing	Increasing
Longitudinal	Increasing	CONSTANT	Increasing
Time-series	Increasing	Decreasing	CONSTANT

Table 2. Types of Perspectives in Which Each Effect is Manifested

Effect present	Perspectives showing significant differences
Aging only	Cross-sectional and Longitudinal
Cohort only	Cross-sectional and Time-series
Time only	Longitudinal and Time-series
Combination of these effects	All three

AGE	Epoch 1 1961-65	Epoch 2 1966-70	Epoch 3 1971-75
80-84	—	—	14
75-79	—	14	18
70-74	14	18	20
65-69	18	20	23
60-64	20	23	40
55-59	23	40	24
50-54	40	24	23
45-49	24	23	18
40-44	23	18	—
35-39	18	—	—

TIME →

Figure 3. Distribution of subjects by age and epoch, $N = 180$. The numbers in the cells are the number of subjects in the specified age-time group.

significant differences as time varies in both these perspectives but is constant in the cross-sectional approach. Finally, it should be noted that if there is a single effect in operation, then it is theoretically impossible for only one perspective to show significant differences as any given effect will manifest itself in two perspectives. Such a situation, however, may occur if, for example, there are large differences in error variances.

If the dependent variable is influenced by more than one of the three independent variables, then all three perspectives should show significant differences. Because their effects are inherently confounded, it will be impossible to assign specific values to each effect. Careful examination of the data may allow one to reach a conclusion about what the main effect(s) is, but such a conclusion would be very tenuous.

Furthermore, whenever one is dealing with several variables, the possibility of combined effects cannot be ignored. The joint effects may simply be additive, or they may be nonadditive, in which case it is said that there is interaction among the variables. This method assumes that there is no interaction among aging, time, and cohort effects; this assumption may lead to some erroneous or oversimplified conclusions.

In the present dietary intake study the data were analyzed with linear regression. For the cross-sectional and longitudinal analyses the slope for the intake of a specified nutrient by age was calculated for each time period and cohort, respectively. For the time series analysis the slope for intake of a specified nutrient by time period was calculated for each age group. A *t* test was then done for the null hypothesis that the slope was zero. A probability level of $p < .05$ was designated as significant in this study.

The results of the analyses will be displayed in an age-time matrix. Figure 3 shows the design that will be employed for each dietary variable, as well as the age and time groupings. Within each cell of the matrix the mean plus or minus the standard error for the specified nutrient is given. The slope obtained in each of the three cross-sectional regressions is shown on the vertical arrow, along with the p value (two-tailed). On the vertical brace is the overall cross-sectional slope and its p value. This slope shows the average cross-sectional change in the dependent variable with an increase of 1 year of age, irrespective of time of observation. The slopes obtained for each of the eight cohorts are shown on the diagonal arrows, along with their p values. The overall longitudinal slope, or the average change among our subjects per year of observation, is given on the diagonal brace. The time series analysis included only the six age groups with values in all three epochs (age groups 45 to 49 through 70 to 74). The slopes for each of these age groups as a function of time of observation are shown on the horizontal arrows, along with their p values. The overall slope for all six age groups is shown on the horizontal brace. This slope shows the average change per year of observation regardless of age. Significant p values are marked with an asterisk.

The slopes for each epoch, each age group, and each cohort were symbolized as b, the overall slopes were called $\bar{b}$. The epoch and age group slopes were calculated by the standard formula for slope, with observations summed over age within the specified epoch (cross-sectional approach) and summed over epoch within the specified age group (time-series approach). The slope for a specified cohort (longitudinal approach) is the mean of the longitudinal slopes calculated for each person within that cohort by the standard formula. The overall cross-sectional and time series slopes were also calculated by the standard formula, with the addition of a summation over epoch in the cross-sectional approach and a summation over age group in the time series. These overall slopes are designated as average slopes in Figure 3. The overall longitudinal slope, however, is designated as a mean slope. This was done to emphasize the fact that it was calculated by a different method. The overall mean longitudinal slope is the sum of the longitudinal slopes for each individual divided by the total number of subjects.

To decide if a given perspective showed significant differences in the intake of a specified nutrient, the following criteria were examined:

(1) Were the majority of slopes statistically significant? The number of slopes that had to be significant for this criterion to be met was at least: two in the cross-sectional approach (out of a total of three slopes); five in the longitudinal approach (out of a total of eight slopes); and, four in the time-series approach (out of a total of six slopes).

(2) Were the slopes consistent in their direction? The number of slopes that had to be in the same direction for this criterion to be met was: three out of three, or all of the slopes, in the cross-sectional approach; at least seven out of eight in the longitudinal approach; and, at least five out of six in the time-series approach.

(3) Was the overall slope, $\bar{b}$, statistically significant? To be designated as showing significant differences, a given perspective had to satisfy at least two of these three criteria. After identifying the perspectives that showed significant differences, the existence of aging, cohort, and/or time effects was then determined by the rationale presented in Table 2. These criteria will be considered further in the discussion section of this paper.

Dietary Intake Study

The Data

The BLSA is an ongoing investigation that began in 1958. The study group consists of community-dwelling, predominantly white, male volunteers who are generally highly educated and who are mainly in, or have retired from, professional or managerial occupations. The social and demographic characteristics of these men have been described by Stone and Norris (1966). Women were added to the study group in 1978, but no dietary data are available on them at present.

The participants visit the Gerontology Research Center in Baltimore at 1- to 2-year intervals depending upon their age. They are housed at the Center for 2.5 days and take part in an extensive series of clinical, physiological, and psychological tests. The BLSA is solely a research project and does not provide treatment for any medical conditions that might be uncovered during the course of testing. However, the participants' physicians are notified if any new conditions are diagnosed.

Collection of data on the composition of the participants' diet began in 1961 and ended in 1975. The method of data collection was a 7-day dietary diary. Registered dieticians instructed the men in the keeping of the diary and used food models to show them how to assess portion size but did not provide dietary counselling. As a learning experience, a trial diary was kept during their stay at the Center and reviewed with a dietician. The men were instructed to record everything eaten during their first "normal" week after returning home. The completed 7-day dietary record was mailed back to the dieticians, who carefully reviewed each record. The men were contacted when questions about specific entries arose. Records were then coded into food lists for computer analysis.

A computer program based on the food lists in USDA Handbook 8 (Watt & Merrill, 1963) was used to convert food intakes to nutrient intakes. This program was developed in the Heart Disease Control Program, Public Health Service, U.S. DHEW, by the late Ms. Olive Hayes and is the program used in the report of McGandy et al. (1966).

A variety of nutrients and dietary variables were examined in the BLSA. Those analyzed in this paper are limited to total calories and calories per kg of body weight; intake of protein, carbohydrate, and fat, and the percentage of calories derived from each of these nutrients; intake of SFA and PFA and the ratio of PFA to SFA intakes; and the amount of cholesterol in the diet. (Other nutrients on which data were collected include fiber, calcium, iron, certain types of vitamins, and alcohol. Data were also collected on the specific type of unsaturated fatty acid that was ingested and on the intake of simple versus complex carbohydrates.)

The 15-year time span in which dietary diaries were completed (1961 to 1975) was divided into three 5-year periods designated as epochs. Data on 180 men with a record in all three epochs were analyzed. Only three of these men were nonwhites. Figure 4 shows the age distribution of the men using the age-time matrix described in the section on research methodology. The numbers in the cells are the number of men in the specified age-time group. The men ranged in age from 35 to 74 when they completed their epoch 1 diary and from 45 to 84 at the time of their epoch 3 diary. Note that all men of a particular initial age move into the next age group in the subsequent epoch; that is, initial age-pentads or cohorts move diagonally in this age-time matrix. The size of the eight cohorts ranged from 14 to 40.

RESULTS

Calories. — Figure 5A shows the mean daily intake of calories by age and epoch. In each epoch a negative slope was obtained; that is, caloric intake decreased with increasing age. The slope was significant in two of the three epochs, and $\bar{b}$ was also significant. This is in accord with our earlier cross-sectional study which showed that caloric intake declined with age parallel to the decrease in basal oxygen consumption and the reduction in physical activity (McGandy et al., 1966). Longitudinally, negative slopes were obtained in seven of the eight cohorts, and were significant in five. The overall slope was also significant. In the time series analysis, five of the six age groups had a decrease in caloric intake between epoch 1 and epoch 3, but the negative slope was significant in only one age group. Thus, following the rationale summarized in Table 2, the data are consistent with a pure aging effect.

Calories per kg of body weight. — The results obtained in the analysis of mean daily intake of calories per kg of body weight generally are the same as those obtained in the analysis of caloric intake unadjusted for weight (Figure 5B). Significant p values, however, were not obtained as frequently. The only overall slope that was statistically significant was that obtained in the longitudinal analysis, but individual cohort slopes were significant only for the three youngest cohorts. The four oldest cohorts lost weight as they aged though (Figure 6), so it is not surprising that their decrease in total caloric intake did not result in significant decreases in caloric intake per kg of weight.

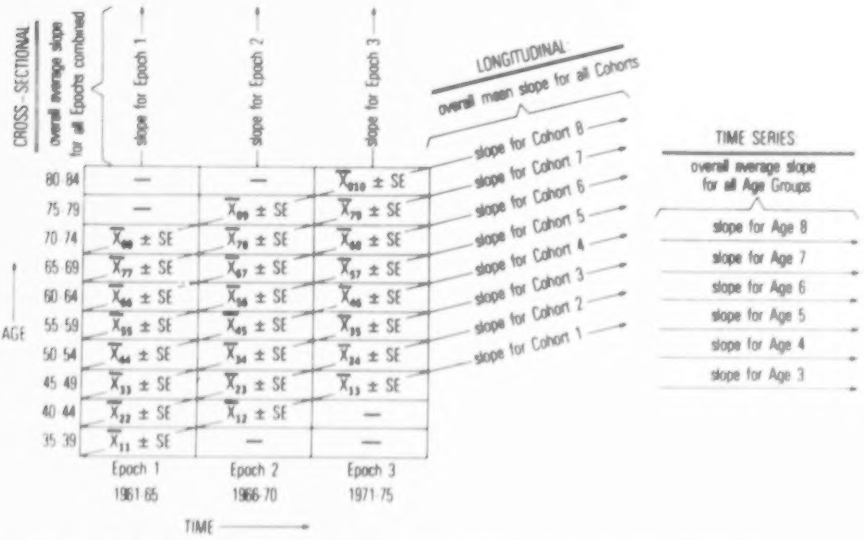

CROSS-SECTIONAL
overall average slope for all Epochs combined
slope for Epoch 1
slope for Epoch 2
slope for Epoch 3
LONGITUDINAL
overall mean slope for all Cohorts
slope for Cohort 8
slope for Cohort 7
slope for Cohort 6
slope for Cohort 5
slope for Cohort 4
slope for Cohort 3
slope for Cohort 2
slope for Cohort 1
TIME SERIES
overall average slope for all Age Groups
slope for Age 8
slope for Age 7
slope for Age 6
slope for Age 5
slope for Age 4
slope for Age 3
AGE
80-84
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
$\bar{X}_{810}$ ± SE
$\bar{X}_{89}$ ± SE
$\bar{X}_{79}$ ± SE
$\bar{X}_{88}$ ± SE
$\bar{X}_{78}$ ± SE
$\bar{X}_{68}$ ± SE
$\bar{X}_{77}$ ± SE
$\bar{X}_{67}$ ± SE
$\bar{X}_{57}$ ± SE
$\bar{X}_{66}$ ± SE
$\bar{X}_{56}$ ± SE
$\bar{X}_{46}$ ± SE
$\bar{X}_{55}$ ± SE
$\bar{X}_{45}$ ± SE
$\bar{X}_{35}$ ± SE
$\bar{X}_{44}$ ± SE
$\bar{X}_{34}$ ± SE
$\bar{X}_{24}$ ± SE
$\bar{X}_{33}$ ± SE
$\bar{X}_{23}$ ± SE
$\bar{X}_{13}$ ± SE
$\bar{X}_{22}$ ± SE
$\bar{X}_{12}$ ± SE
$\bar{X}_{11}$ ± SE
Epoch 1
1961-65
Epoch 2
1966-70
Epoch 3
1971-75
TIME

Figure 4. Design of matrix showing results, where X_{ij} is mean value for cohort i at age j in specified epoch k (i = 1, . . . , 8; j = 1, . . . , 10; and, k = 1,2,3). S.E. = standard error of the mean; slope = b = unit change in specified nutrition variable per year (see text for method of calculation).

Protein. — Mean daily intake of protein declined with increasing age in all three epochs, but none of the individual epoch slopes was statistically significant nor was the mean cross-sectional slope (Figure 5C). Longitudinally, intake of protein declined in all eight cohorts, and in four of them the slope was significant at $p < .05$. The mean longitudinal slope was also statistically significant. Over the time period studied, protein intake declined in five of the six age groups, but the slope was significant in only one of the groups. Thus, there appears to be a small decline in protein intake with increasing age and over time, but the effects are not statistically significant.

Carbohydrate. — Positive cross-sectional slopes were obtained for carbohydrate intake in all three epochs, but the slope was significant in epoch 3 only, and $\bar{b}$ was not significant (Figure 5D). Longitudinally, carbohydrate intake declined in five of the eight cohorts, and all five of the negative slopes were statistically significant, as was $\bar{b}$. In the time series analysis intake of carbohydrates declined in five of the six age groups, but the decline was not significant in any of these five groups. The overall time series slope, however, was statistically significant. These data suggest that there was a negative time effect on carbohydrate intake among our subjects.

Fat. — Cross-sectionally, fat intake decreased with age in all epochs, and the slope was significant in two of the three epochs (Figure 5E). Longitudinally, fat intake decreased in all eight cohorts. The slope was significant in three cohorts and approached statistical significance in two others (with $p = .06$). The mean cross-sectional and mean longitudinal slopes were both significant. Over time, intake of fat declined in five of the six age groups; however, the decline was not significant in any of the five age groups nor in the total time series group. Thus, in contrast to protein and carbohydrate intake, fat intake appears to be influenced by age, intake of this nutrient declining with increasing age. This indicates that the decline in total caloric intake with age is primarily due to a decline in fat intake.

Percent of calories from protein, carbohydrate, and fat. — The relative caloric intake of protein, that is, the percentage of total daily calories derived from this nutrient, showed no aging, cohort, or time effects (Figure 5F). Two of the three cross-sectional slopes were positive, and the other was negative; four of the longitudinal slopes were negative, and four were positive; three of the time-series slopes were positive, and three were negative; all were nearly zero. Furthermore, none of the slopes were statistically significant.

The relative caloric intake of carbohydrates increased with increasing age in all three epochs and all the cross-sectional slopes, albeit small, were statistically significant (Figure 5G). Longitudinally, five of the eight cohorts had positive slopes, but only one was significant. The longitudinal $\bar{b}$ was not significant. In the time series analysis four of the six age groups had negative slopes, but none were statistically significant nor was $\bar{b}$. The cross-sectional data thus suggest that the relative intake of carbohydrates increases with aging, in agreement with the previous observations that the absolute intake of carbohydrates remains constant or increases slightly with age cross-sectionally, while total caloric intake declines. The longitudinal data, however, do not support the presence of such an aging effect. This may be because the longitudinal slopes are also influenced by the effect of time. Whereas the absolute intake of carbohydrates declined over time while total caloric intake remained fairly constant, the longitudinal data on relative intake may have been subject to two opposing effects: an aging effect that acted to increase relative intake and a time effect that acted to decrease it.

In contrast, the relative caloric intake of fat decreased with increasing age in all three epochs, and all the epoch slopes were significant (Figure 5H). Of the eight cohorts, seven had negative slopes, and two of these, as well as $\bar{b}$, were significant. Over time, four of the six age groups had an increase in relative caloric intake of fat, but none of the slopes had significant p values. These data show that the relative caloric intake of fat decreases with age. This is consistent with the aging effect shown for absolute fat intake and implies that the decrease in absolute fat intake with age is proportionately larger than the decrease in total calories with age.

Figure 7 shows the relative caloric intake of each of these nutrients by age with the epochs combined. The percentage of calories derived from protein was constant, approximately 16% in all age groups. The percentage derived from fat declined from approximately 42% in the younger groups to 38% in the older groups. This decrease was compensated by an increase in the percentage of calories from carbohydrates, which rose from roughly 42% in the younger groups to 46% in the older ones. However, as stated above, the latter change was not significant by the criteria employed in this analysis.

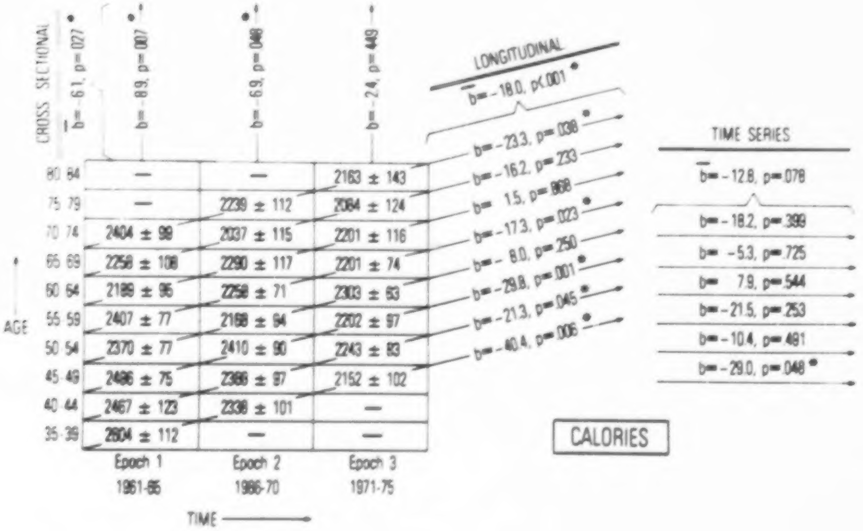


Figure 5A. Mean daily intake of calories.

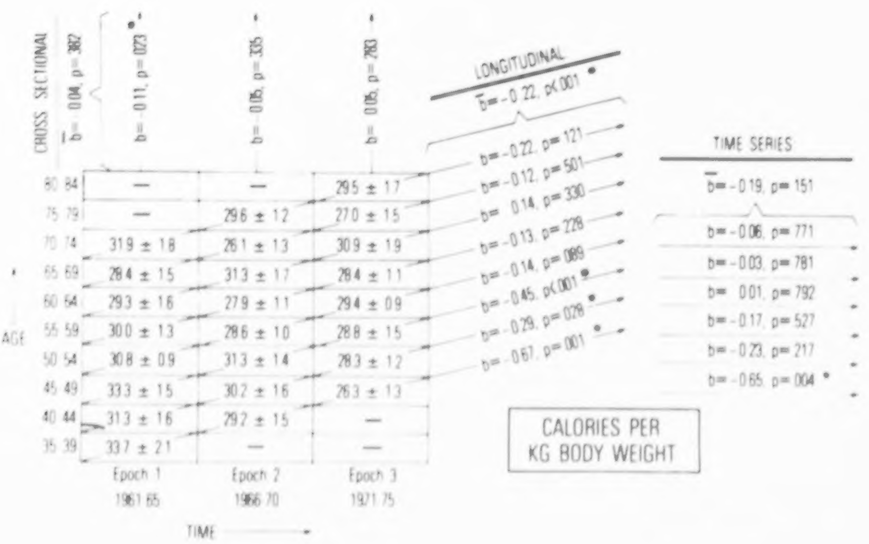
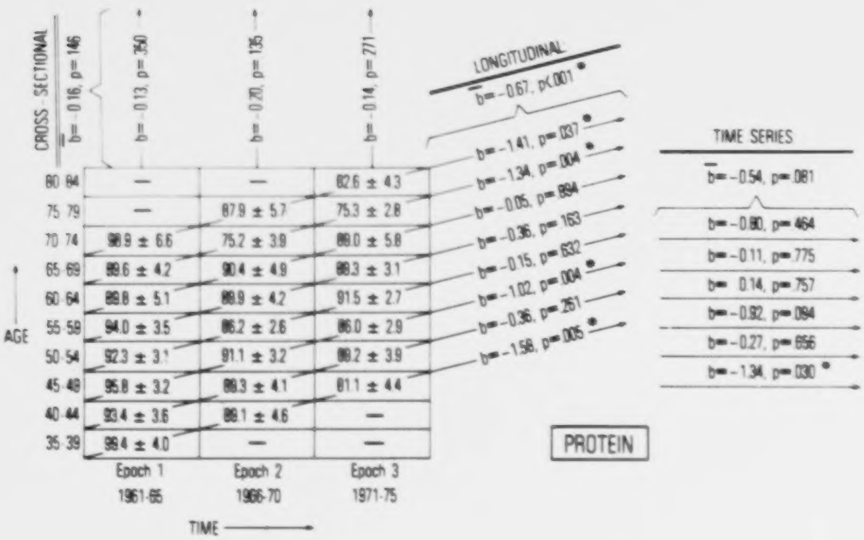
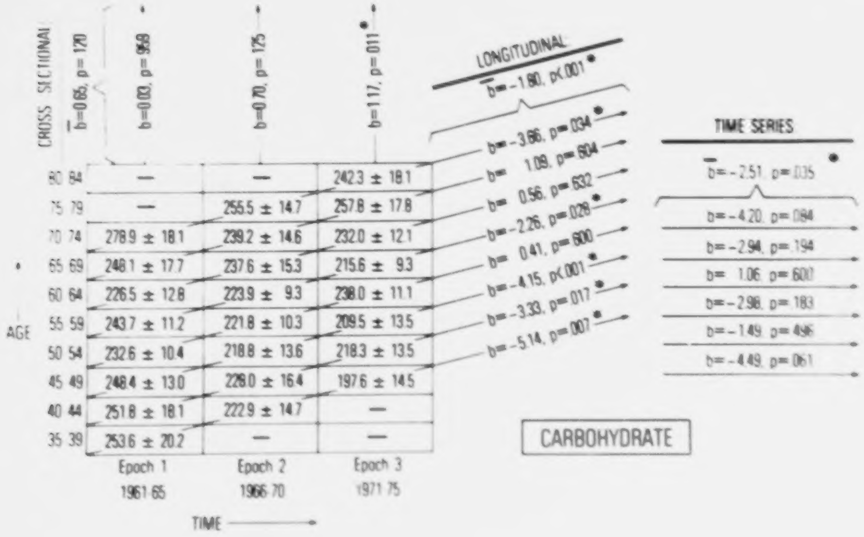


Figure 5B. Mean daily intake of catlories per kg body weight.



AGE	Epoch 1 1961-65	Epoch 2 1966-70	Epoch 3 1971-75
80 84	—	—	82.6 ± 4.3
75 79	—	87.9 ± 5.7	75.3 ± 2.8
70 74	98.9 ± 6.6	75.2 ± 3.9	88.0 ± 5.8
65 69	89.6 ± 4.2	90.4 ± 4.9	88.3 ± 3.1
60 64	88.8 ± 5.1	88.9 ± 4.2	91.5 ± 2.7
55 59	94.0 ± 3.5	86.2 ± 2.6	86.0 ± 2.9
50 54	92.3 ± 3.1	91.1 ± 3.2	88.2 ± 3.9
45 49	95.8 ± 3.2	88.3 ± 4.1	81.1 ± 4.4
40 44	93.4 ± 3.6	88.1 ± 4.6	—
35 39	98.4 ± 4.0	—	—

Figure 5C. Mean daily intake of protein (g)



AGE	Epoch 1 1961 65	Epoch 2 1966 70	Epoch 3 1971 75
80 84	—	—	242.3 ± 18.1
75 79	—	255.5 ± 14.7	257.8 ± 17.8
70 74	278.9 ± 18.1	239.2 ± 14.6	232.0 ± 12.1
65 69	248.1 ± 17.7	237.6 ± 15.3	215.6 ± 9.3
60 64	226.5 ± 12.8	223.9 ± 9.3	238.0 ± 11.1
55 59	243.7 ± 11.2	221.8 ± 10.3	209.5 ± 13.5
50 54	232.6 ± 10.4	218.8 ± 13.6	218.3 ± 13.5
45 49	248.4 ± 13.0	228.0 ± 16.4	197.6 ± 14.5
40 44	251.8 ± 18.1	222.9 ± 14.7	—
35 39	253.6 ± 20.2	—	—

Figure 5D. Mean daily intake of carbohydrate (g)

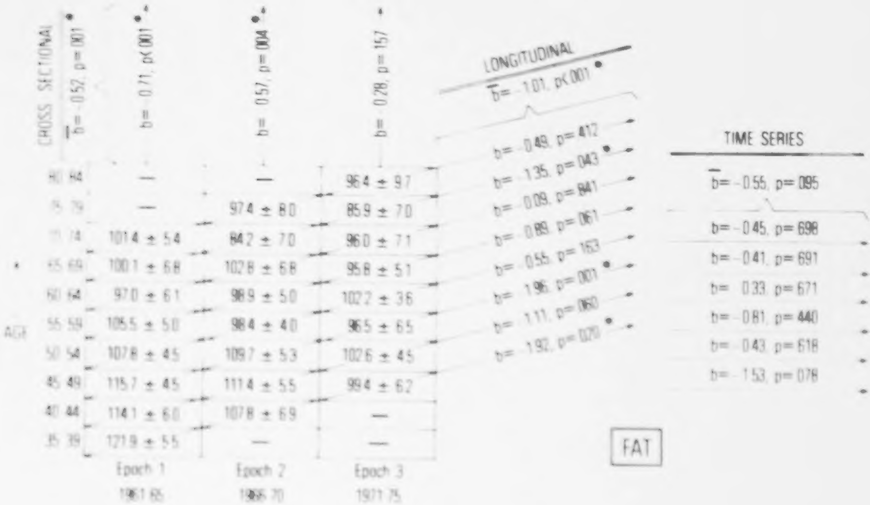

CROSS SECTIONAL
b= -0.52, p=.001 *
b= -0.71, p<.001 *
b= -0.57, p=.004 *
b= -0.28, p=.157
LONGITUDINAL
b= -1.01, p<.001 *
b= -0.49, p=.412
b= -1.35, p=.043 *
b= -0.09, p=.841
b= -0.89, p=.061
b= -0.55, p=.163
b= -1.96, p=.001 *
b= -1.11, p=.060
b= -1.92, p=.020 *
TIME SERIES
b= -0.55, p=.095
b= -0.45, p=.698
b= -0.41, p=.691
b= 0.33, p=.671
b= -0.81, p=.440
b= -0.43, p=.618
b= -1.53, p=.078
AGE
80-84 — — 96.4 ± 9.7
75-79 — 97.4 ± 8.0 85.9 ± 7.0
70-74 101.4 ± 5.4 84.2 ± 7.0 96.0 ± 7.1
65-69 100.1 ± 6.8 102.8 ± 6.8 95.8 ± 5.1
60-64 97.0 ± 6.1 98.9 ± 5.0 102.2 ± 3.6
55-59 105.5 ± 5.0 98.4 ± 4.0 96.5 ± 6.5
50-54 107.8 ± 4.5 109.7 ± 5.3 102.6 ± 4.5
45-49 115.7 ± 4.5 111.4 ± 5.5 99.4 ± 6.2
40-44 114.1 ± 6.0 107.8 ± 6.9 —
35-39 121.9 ± 5.5 — —
Epoch 1 1961-65
Epoch 2 1966-70
Epoch 3 1971-75
TIME
FAT

Figure 5E. Mean daily intake of fat (g).

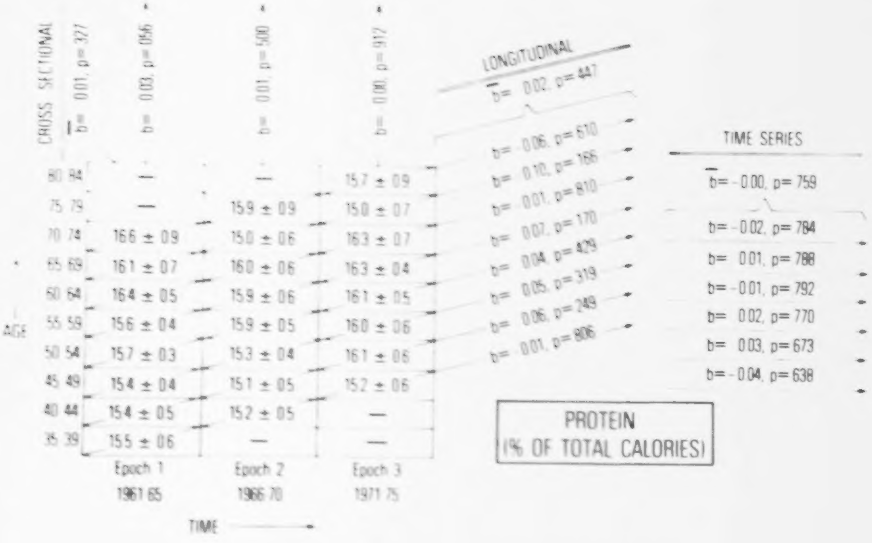

CROSS SECTIONAL
b= 0.01, p=.327
b= 0.03, p=.056
b= 0.01, p=.500
b= 0.00, p=.912
LONGITUDINAL
b= 0.02, p=.447
b= -0.06, p=.610
b= 0.10, p=.166
b= 0.01, p=.810
b= 0.07, p=.170
b= 0.04, p=.429
b= 0.05, p=.319
b= 0.06, p=.249
b= 0.01, p=.806
TIME SERIES
b= -0.00, p=.759
b= -0.02, p=.784
b= 0.01, p=.788
b= -0.01, p=.792
b= 0.02, p=.770
b= 0.03, p=.673
b= -0.04, p=.638
AGE
80-84 — — 15.7 ± 0.9
75-79 — 15.9 ± 0.9 15.0 ± 0.7
70-74 16.6 ± 0.9 15.0 ± 0.6 16.3 ± 0.7
65-69 16.1 ± 0.7 16.0 ± 0.6 16.3 ± 0.4
60-64 16.4 ± 0.5 15.9 ± 0.6 16.1 ± 0.5
55-59 15.6 ± 0.4 15.9 ± 0.5 16.0 ± 0.6
50-54 15.7 ± 0.3 15.3 ± 0.4 16.1 ± 0.6
45-49 15.4 ± 0.4 15.1 ± 0.5 15.2 ± 0.6
40-44 15.4 ± 0.5 15.2 ± 0.5 —
35-39 15.5 ± 0.6 — —
Epoch 1 1961-65
Epoch 2 1966-70
Epoch 3 1971-75
TIME
PROTEIN (% OF TOTAL CALORIES)

Figure 5F. Mean percent of total daily calories derived from protein.

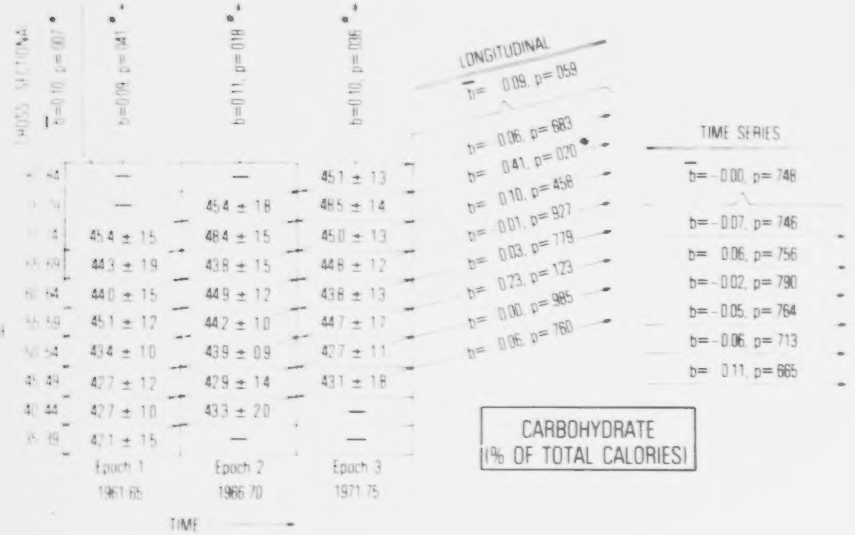


Figure 5G. Mean percent of total daily calories derived from carbohydrate

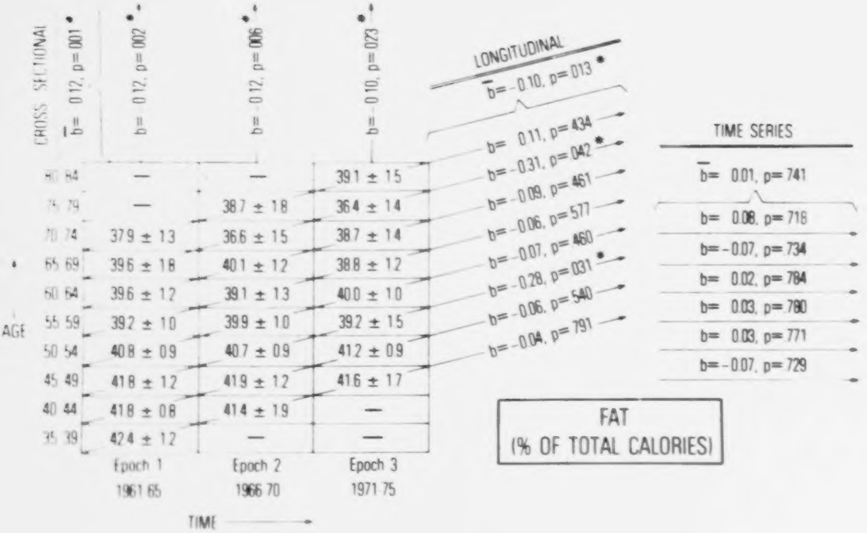


Figure 5H. Mean percent of total daily calories derived from fat

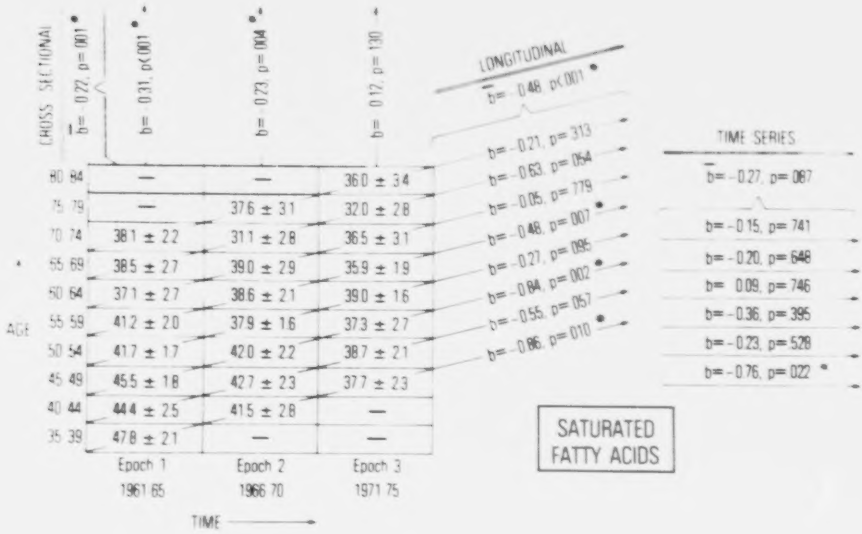

CROSS SECTIONAL
b̄= -0.22, p=.001 *
b= -0.31, p<.001 *
b= -0.23, p=.004 *
b= -0.12, p=.130
LONGITUDINAL
b̄= -0.48, p<.001 *
b= -0.21, p=.313
b= -0.63, p=.054
b= -0.05, p=.779
b= -0.48, p=.007 *
b= -0.27, p=.095
b= -0.84, p=.002 *
b= -0.55, p=.057
b= -0.86, p=.010 *
TIME SERIES
b̄= -0.27, p=.087
b= -0.15, p=.741
b= -0.20, p=.648
b= 0.09, p=.746
b= -0.36, p=.395
b= -0.23, p=.528
b= -0.76, p=.022 *
AGE
80-84 — — 36.0 ± 3.4
75-79 — 37.6 ± 3.1 32.0 ± 2.8
70-74 38.1 ± 2.2 31.1 ± 2.8 36.5 ± 3.1
65-69 38.5 ± 2.7 39.0 ± 2.9 35.9 ± 1.9
60-64 37.1 ± 2.7 38.6 ± 2.1 39.0 ± 1.6
55-59 41.2 ± 2.0 37.9 ± 1.6 37.3 ± 2.7
50-54 41.7 ± 1.7 42.0 ± 2.2 38.7 ± 2.1
45-49 45.5 ± 1.8 42.7 ± 2.3 37.7 ± 2.3
40-44 44.4 ± 2.5 41.5 ± 2.8 —
35-39 47.8 ± 2.1 — —
Epoch 1 1961-65
Epoch 2 1966-70
Epoch 3 1971-75
TIME
SATURATED FATTY ACIDS

Figure 5I. Mean daily intake of saturated fatty acids (g)

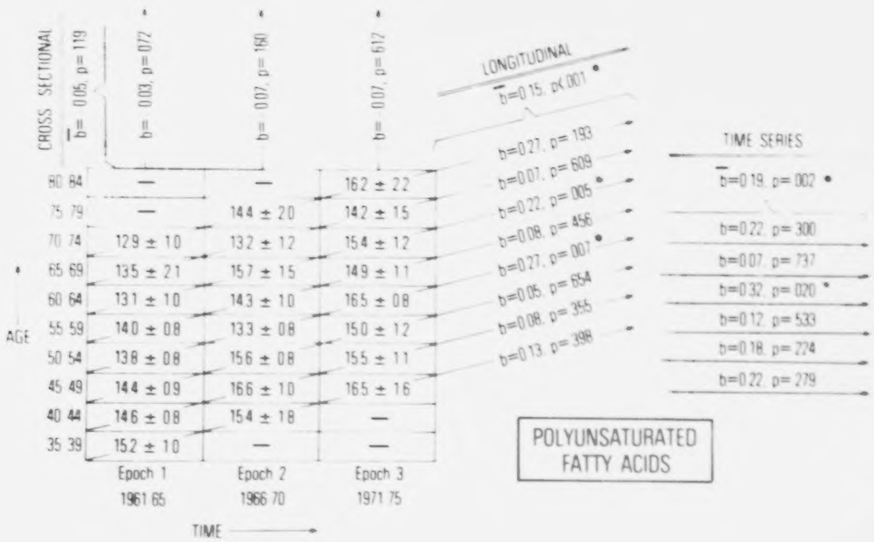
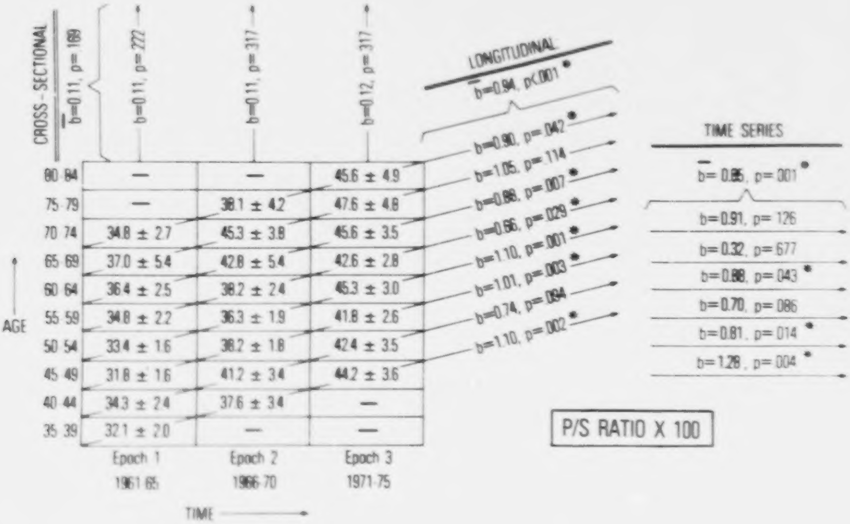
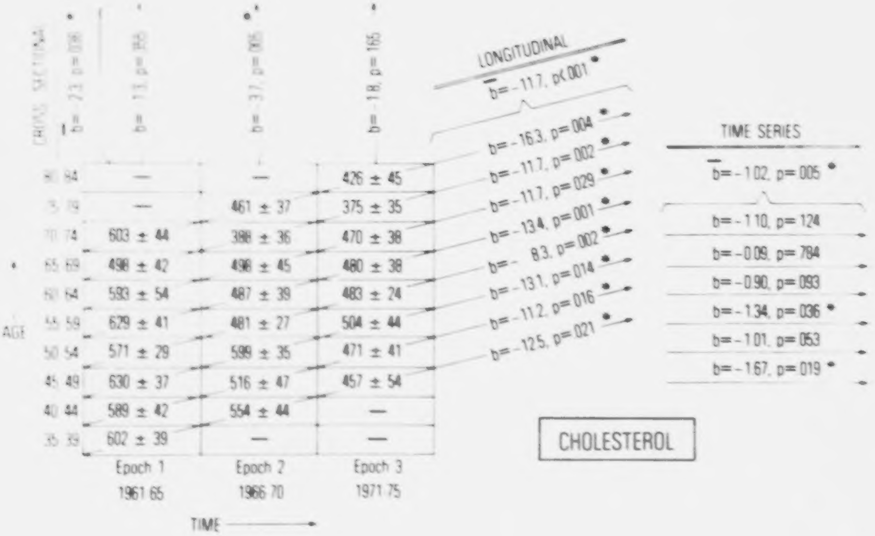

CROSS SECTIONAL
b̄= 0.05, p=.119
b= 0.03, p=.072
b= 0.07, p=.180
b= 0.07, p=.612
LONGITUDINAL
b̄=0.15, p<.001 *
b=0.27, p=.193
b=0.07, p=.609
b=0.22, p=.005 *
b=0.08, p=.456
b=0.27, p=.007 *
b=0.05, p=.654
b=0.08, p=.355
b=0.13, p=.398
TIME SERIES
b̄=0.19, p=.002 *
b=0.22, p=.300
b=0.07, p=.737
b=0.32, p=.020 *
b=0.12, p=.533
b=0.18, p=.224
b=0.22, p=.279
AGE
80-84 — — 16.2 ± 2.2
75-79 — 14.4 ± 2.0 14.2 ± 1.5
70-74 12.9 ± 1.0 13.2 ± 1.2 15.4 ± 1.2
65-69 13.5 ± 2.1 15.7 ± 1.5 14.9 ± 1.1
60-64 13.1 ± 1.0 14.3 ± 1.0 16.5 ± 0.8
55-59 14.0 ± 0.8 13.3 ± 0.8 15.0 ± 1.2
50-54 13.8 ± 0.8 15.6 ± 0.8 15.5 ± 1.1
45-49 14.4 ± 0.9 16.6 ± 1.0 16.5 ± 1.6
40-44 14.6 ± 0.8 15.4 ± 1.8 —
35-39 15.2 ± 1.0 — —
Epoch 1 1961-65
Epoch 2 1966-70
Epoch 3 1971-75
TIME
POLYUNSATURATED FATTY ACIDS

Figure 5J. Mean daily intake of polyunsaturated fatty acids (g)



AGE	Epoch 1 1961-65	Epoch 2 1966-70	Epoch 3 1971-75
80-84	—	—	45.6 ± 4.9
75-79	—	38.1 ± 4.2	47.6 ± 4.8
70-74	34.8 ± 2.7	45.3 ± 3.8	45.6 ± 3.5
65-69	37.0 ± 5.4	42.8 ± 5.4	42.6 ± 2.8
60-64	36.4 ± 2.5	38.2 ± 2.4	45.3 ± 3.0
55-59	34.8 ± 2.2	36.3 ± 1.9	41.8 ± 2.6
50-54	33.4 ± 1.6	38.2 ± 1.8	42.4 ± 3.5
45-49	31.8 ± 1.6	41.2 ± 3.4	44.2 ± 3.6
40-44	34.3 ± 2.4	37.6 ± 3.4	—
35-39	32.1 ± 2.0	—	—

Figure 5K. Mean daily polyunsaturated/saturated fatty acid ratio times 100



AGE	Epoch 1 1961-65	Epoch 2 1966-70	Epoch 3 1971-75
80-84	—	—	426 ± 45
75-79	—	461 ± 37	375 ± 35
70-74	603 ± 44	388 ± 36	470 ± 38
65-69	498 ± 42	498 ± 45	480 ± 38
60-64	593 ± 54	487 ± 39	483 ± 24
55-59	629 ± 41	481 ± 27	504 ± 44
50-54	571 ± 29	599 ± 35	471 ± 41
45-49	630 ± 37	516 ± 47	457 ± 54
40-44	589 ± 42	554 ± 44	—
35-39	602 ± 39	—	—

Figure 5L. Mean daily intake of cholesterol (mg)

Saturated fatty acids. — Intake of SFA declined with age in all three cross-sectional studies, and the slope was significant in two of the three epochs, as was $\bar{b}$ (Figure 5I). Longitudinally, intake of SFA declined in all eight cohorts. Three of the cohorts had significant slopes, and in two more the slopes approached significance ($p < .06$). The overall longitudinal slope was also statistically significant. Over time, intake of SFA declined in five of the six age groups, but the decline was significant in only one group. Thus, there appears to be a decrease in SFA intake with increasing age and a small, but not statistically significant, decrease over time.

Polyunsaturated fatty acids. — Intake of PFA declined with age in all three epochs, but none of the cross-sectional slopes were statistically significant, nor was $\bar{b}$ (Figure 5J). Longitudinally, in contrast to SFA intake, intake of PFA increased in all eight cohorts, and although only two of the longitudinal slopes were statistically significant, the overall slope was significant. In the time series analysis the slope of PFA intake over the three epochs was positive in all six age groups but significant in only one. $\bar{b}$ was, however, statistically significant. These results indicate that intake of PFA has increased over time among our subjects.

P/S ratio. — The ratio of polyunsaturated to saturated fatty acid intake (the P/S ratio) exhibited a time effect, but no aging or cohort effects. This ratio increased cross-sectionally with age in all three epochs, but none of the epoch slopes were statistically significant nor was $\bar{b}$ (Figure 5K). This agrees with the previous observation that intake of both PFA and SFA declined cross-sectionally with age and implies that, although the cross-sectional decline in PFA was not significant, it was sufficiently large to prevent a statistically significant increase in the P/S ratio. In contrast, longitudinally the P/S ratio increased in all eight cohorts, and the slope was significant in six of the cohorts; $\bar{b}$ was also significant. Over time, the ratio increased in all six age groups, and the slope was significant in three of the six groups, as well as overall. Since both the longitudinal and time series perspectives showed significant differences while the cross-sectional perspective did not, the data are consistent with a pure time effect. The P/S ratio increased from a mean of .35 in epoch 1 to .44 in epoch 3 among the age groups in the time series analysis, a 26% increase. This change reflects the increase over time in PFA intake among our participants, and the small, but statistically insignificant, secular decrease in SFA intake.

Cholesterol. — Cross-sectionally, cholesterol intake decreased with age in all epochs, and although the slope was significant only in epoch 2, the overall slope was significant. Longitudinally, intake of this nutrient went down in all eight cohorts, and the cohort slopes were all statistically significant (Figure 5L). In the time series analysis, cholesterol intake decreased in all six age groups, and although the slope was significant in only two

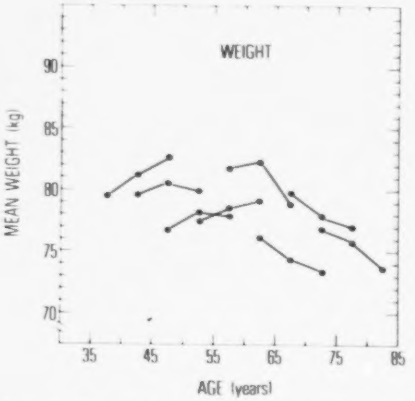


Figure 6. Mean weight (kg) by age and cohort. Each of the eight lines represents an individual cohort.

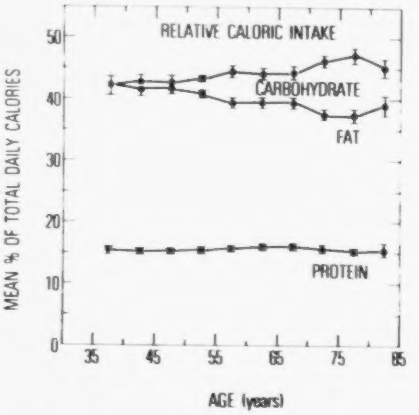


Figure 7. Mean percentage of total daily calories from specified nutrient, by age, epochs combined. Vertical bars represent standard errors of the means.

of them, b was significant. All three perspectives thus showed significant differences in cholesterol intake according to the criteria used here, indicating the presence of multiple effects, but the specific types of effects present cannot be identified. The longitudinal perspective, however, satisfied all three criteria, whereas the cross-sectional and time series did not satisfy criterion 1. (All of the eight cohort slopes were statistically significant, but only one of the three epoch and two of the six age-group slopes were significant.) The longitudinal slopes were also much larger than the cross-sectional or time series slopes. If a cohort effect were one of the main effects present, one would not expect to find that the longitudinal slopes exhibit the most significance, as the cohort variable is constant in this approach. The results may thus reflect the existence of aging and time effects only, both of which would be manifested in significant differences longitudinally. Furthermore, since cholesterol intake appears to decline both with age and over time according to these data, the striking longitudinal differences may reflect the combined impact of aging and time effects. The decline over time may be due in large part to the benefits attributed to a low cholesterol diet by many health professionals in recent years. Among the six age groups in our time series analysis (encompassing ages 45 to 74 years), daily cholesterol intake decreased from a mean of 587 mg in epoch 1 to 478 mg in epoch 3, a 19% decrease.

DISCUSSION

Table 3 summarizes our conclusions about which perspectives showed significant differences and lists the criteria met by each of the "significant" perspectives. To be designated as showing significant differences a perspective had to satisfy at least two of the three criteria presented in the Research Methodology Section.

Half of the 12 nutrients examined exhibited significant cross-sectional differences by the criteria used, and five of these six cross-sectional differences satisfied all three criteria. The greatest number of significant differences was obtained in the longitudinal analyses (10 out of 12), but only three of these met all the criteria. The time series analyses yielded the fewest significant differences (four out of 12), and none of these satisfied all three criteria. These results may reflect several factors. First, given that one slope was statistically significant, criterion 1 (that the majority of slopes were significant) would be met in the cross-sectional perspective if there was only one more significant slope.

Table 3. Conclusions Regarding Which Perspectives Showed Significant Differences, and the Type of Effect Present, by Nutrient

	Perspectives designated significant by Criteria Met[a]			Effect present and its direction	
Nutrient	Cross-sectional	Longitudinal	Time-series	Effect[b]	Direction[c]
Calories	m,d,$\bar{b}$	m,d,$\bar{b}$		Aging	Down
Cals/kg		d,$\bar{b}$			
Protein		d,$\bar{b}$			
Carbohydrate		m,$\bar{b}$	d,$\bar{b}$	Time	Down
Fat	m,d,$\bar{b}$	d,$\bar{b}$		Aging	Down
% Protein					
% CHO	m,d,$\bar{b}$				
% Fat	m,d,$\bar{b}$	d,$\bar{b}$		Aging	Down
SFA	m,d,$\bar{b}$	d,$\bar{b}$		Aging	Down
PFA		d,$\bar{b}$	d,$\bar{b}$	Time	Up
P/S		m,d,$\bar{b}$	d,$\bar{b}$	Time	Up
Cholesterol[d]	d,$\bar{b}$	m,d,$\bar{b}$	d,$\bar{b}$	Multiple	Down

[a] The criteria to be met and the symbols used in this table for these criteria are m (the majority of slopes were statistically significant, $p < .05$), d (the slopes were consistent in their direction) and, $\bar{b}$ (the overall slope $\bar{b}$, was statistically significant, $p < .05$). To be designated as showing significant differences, any given perspective had to satisfy at least two of these three criteria. There is no entry if none or only one of these criteria was met.

[b] Based on the rationale outlined in Table 2. There is no entry if the specified nutrient did not exhibit aging, time, or cohort effects.

[c] The direction indicates the direction of change in the dependent variable as age increases for the aging effects and as period of observation increases for the time effects.

[d] Cholesterol intake exhibited multiple effects. The results were interpreted to primarily reflect a decline in intake with increasing age and over time.

But in the longitudinal perspective there would have to be four more significant slopes, and in the time series three more significant slopes. Secondly, attainment of statistical significance by the longitudinal and time-series slopes was probably hampered by the small size of some of the cohort and age groups. Thirdly, the cross-sectional and time-series designs did not involve repeated measures in the same individuals and thus had larger error variances than the longitudinal design. (This may be why statistically significant differences were found only in the longitudinal perspective for both caloric intake per kg of weight and protein intake.) Finally, the time series analysis was based on 15 years of observation (three epochs), whereas the age range of the men covered 40 years (eight age groups). Botwinick and Arenberg (1976) have shown that the length of the time span and the number of levels of observation can affect statistical significance.

Table 3 also summarizes our conclusions about the types of effects that were present in these dietary data. They were derived by using the rationale given in Table 2 and the data in Figures 5A to 5L. Four of the nutrients exhibited pure aging effects, four exhibited pure time effects, one exhibited multiple effects, and none exhibited cohort effects, based on our definitions of aging, time, and cohort effects. Since the dependent variables in this study are interrelated,the final set of conclusions was evaluated for internal coherence. These conclusions, however, inevitably represent our interpretation of the results.

With the exception of carbohydrate, the absolute intake of all the nutrients studied tended to decline with increasing age. Carbohydrate intake exhibited a very small increase with increasing age. But aging effects were concluded to be significant for only four nutrients. These were calories, fat, SFA, and cholesterol. The percentage of calories derived from fat also declined significantly with increasing age.

Cohort effects were not observed for any of the 12 variables. However, the homogeneity of the BLSA men in terms of their social and economic characteristics may have blurred any differences that could be attributed to the cohorts having experienced different historical events or period effects during their lives. In the general population cohort effects might be seen.

Over the 15-year time period studied there was a slight decline in the absolute intake of protein, carbohydrate, and fat, as well as in the intake of SFA, cholesterol, and total caloric intake. But only the intake of carbohydrate and cholesterol exhibited changes that were consistent with a time effect. The absolute intake of PFA and the P/S ratio increased over time, and both changes were judged to be significant. On the average, the P/S ratio increased by 26% between epoch 1 and 3, and cholesterol intake decreased by 19%. It is likely that these changes reflect the impact of public health efforts to alter the diet of Americans.

Where possible our results were compared with those obtained in the dietary intake studies cited earlier. That is, our cross-sectional results were compared with those obtained in the Framingham Study (Kannel & Gordon, 1970) and HANES I Study (U.S. DHEW, 1977); our longitudinal findings were compared with those of Steinkamp et al. (1965) and Garcia et al. (1975); our time series results were compared with the time trends in the USDA consumption figures. Whereas there are major differences between these studies and ours (such as the instruments used and the nature of the study group) the comparison involved similarities in the direction of changes in nutrient intakes by age or over time rather than specific values of any changes. The Ten-State Nutrition Survey (U.S. DHEW, 1972) was not included in these comparisons. The only adult men who were eligible for that study were those age 60 and over, and their data were not presented by any finer age breakdown.

The Framingham Study used a dietary interview to obtain information on the individual's usual food habits, a method which assumes that people have relatively fixed food patterns over many years (see Mann et al., 1962, for details of the methodology). Data were collected between 1957 and 1960 on a sample of persons returning for their fourth or fifth biennial examination. Persons with a diagnosis of heart disease or hypertension were excluded from the study. Of interest here are the results for 437 men who ranged in age from 37 to 69.

The HANES I Study collected information on a representative probability sample of the total U.S. population aged 1 to 74 years. This survey began in 1971 and was completed in 1974. Diet was assessed by means of a 24-hour recall interview, in which the respondents were asked to report anything they ate or drank during the day preceding the examination (U.S. DHEW, 1973). The BLSA data were compared with the results obtained for 2,562 white men aged 35 to 74 who had incomes above the poverty level.

Our cross-sectional findings generally agreed with those obtained in the Framingham and HANES studies. These two studies, however, did not include all of the 12 dietary variables in the present study. Data are available from the HANES survey on only three of these variables — calories, calories per kg of body weight, and protein intake. The Framingham Study contains data on calories; absolute and relative intake of protein, carbohydrate, and fat; and cholesterol intake.

As in the BLSA, intake of calories declined cross-sectionally with age in both studies, and calories per kg of body weight declined with age in the HANES study. We concluded, however, that among our subjects only total caloric intake exhibited a significant aging effect. Protein intake also declined cross-sectionally with age in all three studies, but our cross-sectional results were not statistically significant. The Framingham Study found that fat intake declined with increasing age, and we concluded that aging did have a significant negative effect on fat intake. In contrast, carbohydrate intake exhibited a small, but statistically insignificant increase with age in our study, while it declined steadily among the Framingham subjects. This discrepancy in results may be due to the different time frames of the BLSA and Framingham studies, to differences in the study groups, or to some unidentified factor(s). The Framingham Study showed no difference with age in the relative caloric intake of protein, carbohydrate, and fat among its participants. The percentage of calories derived from protein was 20%, from carbohydrate 41%, and from fat 39%. These values are of roughly the same magnitude as the BLSA values, but we observed a statistically significant cross-sectional increase with age in the percentage of calories from carbohydrates and a significant decrease in the percentage from fat. Only the relative caloric intake of fat, however, was judged to exhibit a significant aging effect. The final variable included in both the present study and the Framingham Study was cholesterol intake. In the Framingham Study cholesterol intake did not exhibit a consistent relationship with age, while we observed a cross-sectional decline. We concluded that aging did have a negative effect on cholesterol intake, but as this nutrient appeared to be subject to multiple effects, this conclusion is tenuous.

Steinkamp et al. (1965) obtained 1-day dietary records from 141 individuals in each of four study years — 1948, 1952, 1954, and 1962. The participants ranged from age 50 to over age 80 at the start of the study. The authors comment that there was a decrease in total caloric intake with aging, in accord with our findings. They, however, presented longitudinal dietary data for protein intake only. The 68 men in their study showed a .8% decrease in protein intake per year; the men in our study showed a nearly identical rate of decline. Among the 73 women included in their analysis, there was essentially no change in protein intake.

In the study by Garcia et al. (1975), 7-day weighed food intake records were obtained from 28 women during at least two of four study periods — 1948, 1958 to 1960, 1963 to 1966, and 1968 to 1969. The women ranged in age from 29 to 82 years at the time of their first record. By longitudinal analysis, they showed statistically significant decreases in fat and SFA, nonsignificant decreases in calories and carbohydrate, and nonsignificant increases in protein and linoleic acid with aging. In contrast, in our studies in men all of these variables except PFA showed statistically significant decreases longitudinally. The significant increase in PFA in our group may reflect the generally later time period of our study (1961 to 1975) as compared with that of Garcia (1948 to 1969).

The BLSA data on temporal trends were compared with Department of Agriculture figures on changes in food consumption between 1965 (the end of epoch 1) and 1976 (roughly the end of epoch 3). These food consumption figures were derived by, first, determining the quantity of food produced in a specific year; then, making adjustments for stocks on hand at the beginning and end of the year, for food imported and exported during the year, and for food supplies not used for human consumption; and, finally, by reducing these figures to allow for waste between farm and store (Brewster & Jacobson, 1978). Per-person-consumption data were then derived by dividing total consumption by the total U.S. population, with military personnel and their food use excluded. These data are national averages and are not classified by age, sex, socioeconomic status, or other demographic characteristics. USDA consumption figures were available for all but two of the 12 nutrition variables under study here; these two were calories per kg of body weight and cholesterol intake.

We observed a small, but statistically insignificant, decrease in caloric intake in our time series analysis. The USDA data on the other hand showed a slight increase in average caloric intake between 1965 and 1976. Their data, however, are not adjusted for age changes; therefore, the increase in caloric intake may reflect the fact that in this period the "baby boom" generation, (i.e., those born during the decade of high fertility rates following World War II), reached their teenage and young adulthood years, ages associated with high food consumption. Intake of protein did not exhibit any significant changes over time in either the BLSA or USDA data. In contrast, intake of both carbohydrate and fat declined over time in the BLSA group, while intake of both these nutrients increased in the total population according to the USDA figures. Again, the unadjusted nature of the USDA data may account for this discrepancy. We concluded, however, that among our subjects only carbohydrate intake showed a significant time effect. The time trends in the relative caloric intake of protein, carbohydrate, and fat based on the USDA data are in the same direction as the BLSA trends (i.e., no change for protein, a decrease for carbohydrate, and an increase for fat). The time changes among our participants in percentage of calories from carbohydrate and fat, however, were not statistically significant. Intake of SFA declined over time among our subjects, but the time-series results were not statistically significant. The USDA figures show a small increase in SFA consumption between 1965 and 1972, followed by a small decline between 1972 and 1976. The time-series results for PFA were statistically significant, and we concluded that PFA intake decreased over time among our participants. This is in accord with the USDA data which showed that average consumption of PFA increased by roughly one-third between 1965 and 1976. As would be expected, both the BLSA and USDA data show an increase over time in the P/S ratio. In our study the P/S ratio increased from a mean of .35 in epoch 1 to .44 in epoch 3. According to USDA data, in the total U.S. population the average P/S ratio increased from roughly .32 in 1965 to .45 in 1976, values remarkably close to the BLSA figures.

This analysis of the dietary diaries completed by 180 men in the BLSA between 1961 and 1975 has been descriptive in nature. We have also attempted to identify pure aging, cohort, and time effects as we defined such effects. If multiple effects are present they cannot be separated by the approach used here, or any other approach, as they are inherently confounded with each other.

The criteria used to determine if nutrient intake varied significantly in a given perspective were based on our judgment of what was needed to indicate the existence of a significant deviation. We also have not considered the importance of testing, selective survival, and interaction effects. The conclusions reached are thus debatable. Nevertheless, the data comprise a unique set of longitudinal dietary information on a middle- to upper-middle-class group of white men.

In addition to detailed dietary records, many types of physiological, medical, and biochemical data have been collected systematically on these men. Thus, it is possible to evaluate the importance of the aging and time effects that were proposed here. Hershcopf et al. (1982) have already examined the relationship between changes in serum cholesterol levels over the period of 1963 to 1977 among BLSA participants, and changes in dietary constituents. They concluded that diet could not fully explain the observed drop in serum cholesterol and that some as yet undetermined factor(s) was responsible for this change. Additional research areas could include the correlaton between changes in carbohydrate intake as these men age and their performance on glucose tolerance tests, and the association between diet and the development of such diseases as arteriosclerosis and osteoporosis. The longitudinal nature of the BLSA study and the variety of data that are collected will permit epidemiologic analyses of diet and health that are not possible in other types of studies.

REFERENCES

Adam, J. Sequential strategies and the separation of age, cohort, and time-of-measurement contributions to developmental data. *Psychological Bulletin*, 1978, *85*, 1309–1316.

Baltes, P. B. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, 1968, *11*, 145–171.

Botwinick, J., & Arenberg, D. Disparate time spans in sequential studies of aging. *Experimental Aging Research*, 1976, 2, 55–61.

Brewster, L., & Jacobson, M. F. *The changing American diet.* Center for Science in the Public Interest, Washington, DC, 1978.

Garcia, P. A., Battese, G. E., & Brewer, W. D. Longitudinal study of age and cohort influences on dietary patterns. *Journal of Gerontology*, 1975, *30*, 349–356.

Hershcopf, R. J., Elahi, D., Andres, R., Baldwin, H. L., Raizes, G. S., Shocken, D. D., & Tobin, J. D. Longitudinal changes in serum cholesterol in man: An epidemiologic search for an etiology. *Journal of Chronic Diseases*, 1982, *35*, 101–114.

Kannel, W. B., & Gordon, T. The Framingham diet study: Diet and the regulation of serum cholesterol. *The Framingham Study. An epidemiological investigation of cardiovascular disease.* Section 24. U.S. Department of Health, Education and Welfare, Public Health Service, Washington, DC, 1970.

Mann, G. V., Pearson, G., Gordon, T., & Dawber, T. R. Diet and cardiovascular disease in the Framingham Study. I. Measurement of dietary intake. *American Journal of Clinical Nutrition*, 1962, *11*, 200–225.

Mason, K. O., Mason, W. M., Winsborough, H. H., & Poole, W. K. Some methodological issues in cohort analysis of archival data. *American Sociological Review*, 1973, *38*, 242–258.

McGandy, R. B., Barrows, C. H., Jr., Spanias, A., Meredith, A., Stone, J. L., & Norris, A. H. Nutrient intakes and energy expenditure in men of different ages. *Journal of Gerontology*, 1966, *21*, 581–587.

Schaie, K. W. A general model for the study of developmental problems. *Psychological Bulletin*, 1965, *64*, 92–107.

Schlenker, E. D., Feurig, J. S., Stone, L. H., Ohlson, M. A., & Mickelsen, O. Nutrition and health of older people. *The American Journal of Clinical Nutrition*, 1973, *26*, 1111–1119.

Steinkamp, R. C., Cohen, N. L., & Walsh, H. E. Resurveys of an aging population. Fourteen year follow-up. *Journal of the American Dietetic Association*, 1965, *46*, 103–110.

Stone, J. L., & Norris, A. H. Activities and attitudes of participants in the Baltimore Longitudinal Study. *Journal of Gerontology*, 1966, *21*, 575–580.

U. S. Department of Health, Education, and Welfare. *Ten-State Nutrition Survey, 1968–1970. V. Dietary*. U. S. Government Printing Office, Washington, DC, 1972.

U. S. Department of Health, Education, and Welfare. Plan and Operation of the Health and Nutrition Examination Survey, United States — 1971–1973. *Vital and health statistics*, Series 1, *Programs and collection procedures*, No. 10b, 1973.

U. S. Department of Health, Education, and Welfare. Dietary intake findings, United States, 1971–1974. *Vital and health statistics*, Series 11, *Data from the National Health Survey*, No. 202, 1977.

Watt, B. K., & Merrill, A. L. *Composition of foods—raw, processed, prepared* (Rev.). U. S. Department of Agriculture, Agriculture Handbook No. 8, Washington, DC, 1963.

Journal of Gerontology
1983, Vol. 38, No. 3, 307-314

Longitudinal Chest X-Ray Changes in Normal Men[1]

Robert E. Ensor, MD,[2,4] Jerome L. Fleg, MD,[3,4]
Young C. Kim, MD,[2,4] Erlinda F. de Leon, MD,[2,4]
and Stanford M. Goldman, MD[2,4]

To determine the changes on chest x-ray attributable to the aging process, we evaluated cardiovascular and pulmonary structures on two standard postero-anterior chest x-rays taken at least 10 years apart (*M* = 16.9 years) in 67 carefully screened healthy men initially aged 23 to 76 years. The aortic knob diameter increased in 79% of subjects. Although mean cardiothoracic ratio increased overall, only 3% of men developed a cardiothoracic ratio greater than .50, and none exceeded .51. Pulmonary abnormalities on initial chest x-ray consisted mainly of hyperinflation (27%) and increased markings (19%), both of which doubled in prevalence during follow-up. Kerley B lines and enlarged pulmonary arteries were rare initially but increased three- to five-fold. The prevalence of these findings did not differ between smokers and nonsmokers. Based on commonly accepted x-ray criteria, chronic obstructive lung disease was suggested in 15% of the initial films and 21% of the final films despite the absence of clinical or spirometric abnormalities.

Key Words: Aging, Aorta, Heart, Lung

DESPITE the ubiquity of the standard chest roentgenogram in clinical medicine, there has been little systematic evaluation of what constitutes the normal changes with age on this examination. Previous studies of x-ray findings in normal persons have consisted of evaluating apparently healthy individuals being seen for noncardiopulmonary disorders (Cowan, 1959; Edge et al., 1964; Felson, 1973; Lauder & Milne, 1976). Extensive clinical evaluations of their cardiopulmonary status were not performed routinely. Because of this lack of detailed screening in selecting these samples, the effects of age versus unrecognized cardiopulmonary disease were probably not separated adequately. These studies were usually cross-sectional in their approach (i.e., they analyzed chest x-rays performed on only a single occasion).

In seeking to determine the normal aging-induced changes in the cardiovascular and pulmonary structures on a standard chest x-ray, the present study is unique in several aspects. The study population was derived from male volunteers in a longitudinal study on aging. In order to qualify as "normal," a man initially had to be free of cardiovascular and pulmonary disease as established by rigorous clinical and laboratory testing and had to remain so on biennial visits throughout the follow-up period that averaged 16.9 years.

METHODS AND MATERIALS

Between the years 1958 and 1979, 1,115 male volunteers were enrolled in the Baltimore Longitudinal Study on Aging (Stone & Norris, 1966). These men underwent 2.5 days of extensive clinical evaluation biennially at the National Institute on Aging's Gerontology Research Center. A 1.83m standard postero-anterior chest x-ray was performed on each man during every visit. From this pool of participants, we selected those who were followed for at least 10 years. In addition, each of our subjects had to fulfill the following selection criteria both initially and during every subsequent visit throughout the study:

(1) Absence of systemic disease based upon history and physical examination and routine laboratory tests (complete blood count, creatinine, glucose tolerance test, urinalysis)

---

[1] These results were presented at the 1981 Annual Meeting of University Radiologists. We wish to thank Dr. Edward G. Lakatta for his critical review of the manuscript, Ms. Frances Buscemi for her clerical help, Mrs. Elsie Beard for statistical analyses, and Mrs. Martha M. Wolfe and Ms. Joyce L. Schaefer for their secretarial assistance. Requests for reprints should be addressed to Dr. Stanford M. Goldman, Department of Radiology, Baltimore City Hospitals, 4940 Eastern Avenue, Baltimore, MD 21224.

[2] Department of Radiology, Baltimore City Hospitals, Baltimore, MD.

[3] Cardiovascular Division, Gerontology Research Center, National Institute on Aging, Baltimore, MD.

[4] Johns Hopkins University School of Medicine, Baltimore, MD.

(2) Normal resting electrocardiogram and treadmill exercise test (double Master tests were performed prior to 1969).

(3) No medication affecting cardiovascular or pulmonary function.

(4) Forced vital capacity (FVC) within one standard error of the estimate of the values predicted by age and height (Cardiopulmonary Council of the American Heart Association, 1982) and ratio of 1-sec forced expiratory volume (FEV_1) to FVC within one standard deviation of age-adjusted normal values throughout the study.

(5) Blood pressure less than 140/90 throughout the study.

Each man's initial and most recent chest roentgenograms were evaluated independently by two board-certified radiologists after all identifying markings were removed. Afterwards, any disagreement in the interpretation was resolved by consensus. In the first 15 cases all 40 parameters described by Felson (1973) were used. It was noted that only 12 of these actually changed during the period of observation. Therefore, only these measurements were repeated on the remaining 52 men. Additionally, the height of the dome of the right hemidiaphragm was measured as described by Nicklaus et al. (1966) and by Krumpelman and Burki (1980). The status of the peripheral lung vasculature as reported by Reid and Millard (1964) was also ascertained. The criteria used for determining chronic obstructing pulmonary disease (COPD) are shown in Table 1.

A numerical value was given to each abnormality on a scale of 0 to 3, 0 being normal and 3 being severe. The five scales were then summed. Radiographic evidence of COPD was defined by a total score of eight or more based on this scoring system.

Longitudinal changes were analyzed by paired *t* test; cross-sectional comparisons were made with the unpaired *t* test. Among groups with initial ages 20 to 39 ($n = 19$), 40 to 59 ($n = 35$), and 60 or older ($n = 13$), cross-sectional comparisons were made using one-way analysis of variance. The independent effects of weight changes and age on longitudinal cardiovascular changes were assessed by multiple regression analysis. Differences between percentages of men within each of two groups were assessed by chi-square analyses. A p value $\leq .05$ was considered significant. Data are expressed as mean plus or minus standard deviation ($M \pm SD$). Pulmonary findings were compared in smokers and nonsmokers, smoking being defined by the consumption of at least 10 cigarettes daily for 5 or more years.

RESULTS

At the onset of the study most of the 67 men were in their 30s to 50s. Their ages ranged from 23 to 76

Table 1. Criteria for Radiographic Determination of Chronic Obstructive Pulmonary Disease

Signs	Grade		
	I	II	III
Hyperinflation[a]	Localized Minimal flattening of hemidiaphragms	More generalized, bullae Moderate flattening of hemidiaphragms (to level of anterior sixth rib or below)	Generalized Marked flattening of hemidiaphragms (to level of seventh anterior rib[b] or below and diaphragmatic curvature less than 1.5 cm[b] or concave
Diminished arterial vascularity[c]	Mild (often localized)	Moderate	Absent vessels
Increased interstitial markings[d]	Minimal	Moderate	Marked
Enlarged central pulmonary arteries	Minimal (16-18 mm)	Moderate (18-20 mm)	Marked (> 20 mm)
Right ventricular hypertrophy[e]	Mild	Moderate	Severe

[a] Fraser et al., 1979

[b] Nicklaus et al., 1966

[c] Thurlbeck, 1980

[d] Thurlbeck and Simon, 1978

[e] Lusted and Keats, 1967

years (48.0 ± 13.1 years). By the conclusion of the observation period, the majority of the men were in their 50s to 70s, ranging from 37 to 90 years (64.9 ± 12.4 years). Follow-up duration averaged 16.9 ± 3.0 years (range 10 to 22 years) and did not differ significantly among the three age groups (M = 17.6, 17.1 and 15.1 years, respectively in the young, middle aged, and elderly groups). Men initially 20 to 39 years of age gained 2.6 ± 6.4 kg during the study period; men initially 40 to 59 years gained .6 ± 6.2 kg, and men 60 or older lost 7.9 ± 6.3 kg, $F(2,64) = 11.99$, $p < .0001$, $\omega^2 = .247$.

CARDIOVASCULAR FINDINGS

Heart. — No man had a cardiothoracic ratio (CTR) exceeding .50 on initial examination, and only 3% developed this abnormality on the final chest film, the largest CTR measuring only .51. Valvular or coronary calcifications were not detected in any man on either film. Overall, the CTR increased in 55%, decreased in 19%, and remained unchanged in 22% of the men. CTR increased from .407 ± .04 to .426 ± .04 over the study period, $t(65) = 4.87$, $p < .001$, $\omega^2 = .145$. In 13% of the men, the CTR changed by more than .06, increasing in each case.

Only 10% of cardiac transverse diameters increased by more than 2.0 cm between examinations, the largest increase being 3.0 cm. The mean longitudinal increase was .72 ± .99 cm. Initial and final heart sizes for the group were 12.7 ± 1.3 cm and 13.4 ± 1.5 cm, respectively, $t(65) = 5.92$, $p < .001$, $\omega^2 = .202$. Overall, heart size increased in 69% of men, decreased in 13%, and did not change in 18%, closely paralleling the CTR changes. There were no differences in mean age or follow-up duration between the groups whose cardiac dimension or CTR increased by over 2.0 cm or .06, respectively, and those whose did not. Transverse chest diameter increased in 57%, decreased in 22%, and remained unchanged in 21%. The corresponding mean values were 31.1 ± 1.7 cm and 31.4 ± 1.8 cm, $t(65) = 2.98$, $p < .01$, $\omega^2 = .056$ for initial and final films, the mean longitudinal increment being .32 ± .87 cm. Intraobserver variations for CTR, heart size, and chest diameter respectively were 1.3%, .6%, and 1.0%; the corresponding values for interobserver variations were 2.8%, 2.1%, and .9%.

We further characterized the changes in heart size, transverse chest diameter, and CTR according to a man's age at entry into the study (Table 2). Cross-sectionally, both cardiac size and CTR increased with age, although statistical significance was achieved only for the initial visit. Longitudinal increases in heart size and CTR occurred within all three age groups but did not reach significance in the elderly group. No differences among age groups were found in the magnitude of these longitudinal changes adjusted for follow-up duration (Δ year). Transverse diameter did not differ by cross-sectional analysis across the three groups; Δ year for this variable did, however, differ significantly among them due to a tendency for chest size to diminish with time in the elderly men rather than enlarge as in the two younger groups.

Because age-related changes in the anteroposterior (AP) diameter of the chest could theoretically create differences in magnification of thoracic structures solely due to differences in their distance from the x-ray tube, we compared the AP chest diameter, measured anthropomorphically on initial and final visits, as shown in Table 4. Both cross-sectionally and longitudinally, AP diameter increased with age except in the elderly men, in whom this dimension actually diminished slightly over time. For the group as a whole, AP dimension increased from 22.14 ± 1.90 to 23.11 ± 2.41, $t(65) = 4.51$, $p < .001$, $\omega^2 = .126$.

To investigate whether a sizeable change in body weight over the study period might influence the longitudinal change in cardiovascular measurements, we compared men who gained 5.0 kg or more with those who lost a similar amount (Table 3). Longitudinal increments in heart size, chest diameter, and CTR were all significantly greater in the former group despite equal follow-up durations (17.1 and 17.0 years, respectively); it was noted that men gaining weight were 16.7 years younger than their counterparts whose weight declined, 39.2 ± 7.4 years versus 55.9 ± 12.5 years, $t(29) = 4.40$, $p < .001$, $\omega^2 = .372$. By multiple simultaneous regression analysis it was found that initial age and longitudinal weight change (Δ wt) contributed nearly equally to the change in heart size, $F(2.28) = 4.89$, $p < .05$, and $F(2.28) = 5.42$, $p < .05$, respectively. Only age affected the change in chest diameter, $F(2.28) = 8.71$, $p < .01$, accounting for 96% of the variance. Only weight change significantly influenced the change in CTR, $F(2.28) = 4.94$, $p < .05$, accounting for 75% of the variance.

Aorta. — Although the aortic knob diameter enlarged during the study in 79% of men, the increment was 1.0 cm or greater in only 7%; aortic diameter decreased slightly in one man. During the

Table 2. Mean Cross-Sectional and Longitudinal Measurements (in cm) Grouped According to Age at Entry

Measurements	Age Group							
	20-39		40-59		60+		F^a	ω
Heart								
Initial	12.0	(1.0)	12.8	(1.6)	13.3	(.9)	3.70*	.075
Final	12.8	(1.2)	13.7	(1.8)	13.5	(.9)	2.27	.037
Δ/year	.046	(.06)	.053	(.07)	.014	(.05)	2.06	.031
t^b	3.49**		4.90***		1.16			
ω	.227		.247		.013			
Chest								
Initial	30.8	(1.3)	31.2	(1.8)	31.2	(1.8)	.40	.018
Final	31.3	(1.5)	31.6	(1.9)	31.0	(2.1)	.73	.008
Δ/year	.029	(.05)	.025	(.05)	.018	(.05)	3.95*	.081
t	2.80*		2.86**		1.48			
ω	.153		.093		.044			
Cardiothoracic ratio								
Initial	.39	(.03)	.41	(.04)	.43	(.02)	3.82*	.078
Final	.41	(.04)	.43	(.05)	.44	(.03)	2.45	.041
Δ/year	.0010	(.002)	.0013	(.002)	.0008	(.002)	.37	.019
t	2.53*		3.73***		1.83			
ω	.124		.156		.083			
Aorta								
Initial	3.2	(.6)	3.5	(.5)	3.7	(.7)	3.55*	.071
Final	3.6	(.4)	3.8	(.5)	4.0	(.9)	2.03	.030
Δ/year	.025	(.02)	.020	(.02)	.023	(.02)	.42	.018
t	4.65***		6.80***		3.30**			
ω	.352		.393		.276			

Note. Standard deviations are in parentheses.

[a] Cross-sectional comparison.

[b] Longitudinal comparison within age group.

$*p < .05$

$**p < .01$

$***p < .001$

study the proportion of men with aortic knob size exceeding 4.0 cm increased from 12% initially to 28% on final examination. Mean values for aortic size were 3.4 ± .6 cm and 3.8 ± .6 cm, respectively, $t(65) = 8.69$, $p < .0001$, $\omega^2 = .357$, on first and final films; the average longitudinal change was .37 ± .35 cm. Intraobserver and interobserver variation for this measurement was 4.6% and 5.1%, respectively. Aortic dilatation occurred in all three age groups over time; the yearly rate of enlargement (Δ/year) did not differ among them (Table 2). Cross-sectionally, aortic diameter also enlarged with increasing age, though reaching statistical significance only on the initial visit. Weight change during the observation period had no demonstrable effect on the increment in aortic size, in contrast with its effect on other cardiothoracic measurements (Table 3).

The prevalence of calcium deposition in the aortic knob increased from 7% to 12% during the study. Those men demonstrating such calcium deposition on their final chest x-ray were older than those without this finding, 74.3 ± 11.2 years versus 63.6 ± 12.5 years, $t(65) = 2.36$, $p < .01$, $\omega^2 = .064$. Even on final x-ray, however, the degree of calcification was modest in all but one man. Aortic tortuosity increased in prevalence rather dramatically between examinations, from 19% to 61%. Although the men initially showing aortic tortuosity were older than their normal counterparts, 57.2 ± 13.3 years versus 45.8 ± 12.8 years, $t(65) = 2.70$, $p < .01$, $\omega^2 = .086$, no age difference was seen between men who demonstrated such tortuosity on the final chest film and those who did not. The mean aortic diameter did not differ significantly between those men with tor-

Table 3. Effect of Weight Change on Longitudinal Cardiovascular Measurements

Measurement	Weight (kg) Gain (n = 15)[a]		Weight (kg) Loss (n = 16)[b]		t	ω^2
Δ Heart[c]	1.25	(.95)	.16	(.98)	3.08**	.215
Δ Chest	.56	(.65)	.006	(.86)	2.05*	.094
Δ Cardiothoracic ratio	.032	(.027)	.006	(.032)	2.36*	.128
Δ Aorta	.41	(.35)	.29	(.33)	.94	.004
Δ Anteroposterior chest diameter	2.83	(1.16)	.39	(1.46)	6.69***	.585

Note. Entries are means with standard deviations in parentheses.

[a] M = 8.8 ± 3.1, n = 15.

[b] M = −9.7 ± 5.0, n = 16.

[c] Δ = Longitudinal change in cm.

$^{*}p < .05$

$^{**}p < .01$

$^{***}p < .001$

tuous or calcium-containing aortas and those without these abnormalities.

PULMONARY FINDINGS

The prevalence of most of the pulmonary abnormalities tabulated at least doubled over the observation period (Table 5). Similarly, the percentage of men demonstrating at least one of the first five findings listed in Table 5 was 30% initially and increased to 63% by the final visit. Individually, none of these radiographic abnormalities was associated with a cross-sectional age difference from those not demonstrating the abnormality. None of the radiographic pulmonary findings was associated with a percentage of smokers significantly different from the 21% prevalence in the total population.

Although these men were preselected for the absence of clinical or spirometric evidence of obstructive lung disease, 15% of them demonstrated multiple radiographic signs suggestive of COPD on initial visit, increasing to 21% by the conclusion of the study. On both initial and final examinations, men with such x-ray findings were older than those without them (58.6 ± 11.0 years vs. 46.1 ± 13.0 years, $t(65) = 3.08$, $p < .01$, $\omega^2 = .112$) on first visit and 73.2 ± 11.5 years vs. 62.7 ± 12.2 years on final x-ray, $t(65) = 2.91$, $p < .01$, $\omega^2 = .100$. Only 20% (2 of 10) and 14% (2 of 14) of the men with radiographic suspicion of COPD on initial and final films, respectively, were smokers, not different from the prevalence of smokers in the group with normal films.

Although radiographic criteria for COPD are by nature subjective, agreement between the two radiologists who read each film was excellent with regard to detection and classification of COPD. Both radiologists concurred in the categorization of COPD in 100% of the initial x-rays and in 89% of the final films; disagreements were resolved by consensus.

Table 4. Anteroposterior Chest Diameter Grouped According to Age at Entry

Age group	Initial visit	Final visit	t	ω^2
20–39	20.9 (2.31)	22.7 (2.67)	4.30***	.315
40–59	22.5 (1.78)	23.6 (2.24)	4.27***	.198
60+	23.0 (1.71)	22.4 (2.26)	2.24*	.134
F	5.64**	1.58		
ω^2	.122	.017		

Note. Standard deviations are in parentheses.

$^{*}p < .05$

$^{**}p < .01$

$^{***}p < .001$

Table 5. Pulmonary Abnormalities on Initial and Final X-ray

Abnormality	Initial %	Final %
Hyperinflation	27	54
Increased markings	19	42
Diminished arterial vascularity	6	10
Enlarged pulmonary arteries	3	9
Right ventricular hypertrophy	0	0
Kerley B Lines	4	19
Bullae	0	0

DISCUSSION

Our finding that heart size increased with advancing age supports prior data, including a 5-year longitudinal study of the heart size by Lauder and Milne (1976) in a random sample of elderly adults. The increase in heart size in the men in our sample averaged only .7 cm, with 10% of men showing increases of more than 2.0 cm. Whereas none of these men demonstrated clinical cardiovascular disease over a follow-up period averaging more than 17 years, it is doubtful if such an increment in heart size is of any prognostic significance. This increase in cardiac diameter could reflect dilatation of the left ventricle, right ventricle, and/or the right atrium, the structures comprising the cardiac silhouette in the frontal view. The .55% magnification factor resulting from a .97 cm longitudinal increase in AP chest diameter at a standard 6-foot examination distance could account for at most 10% of this increase in heart size with time. M-mode echocardiographic studies (Gerstenblith et al., 1977; Henry et al., 1980) have not suggested age-related ventricular dilatation but have shown a modest concentric hypertrophy of the left ventricular wall that could account for perhaps 25% of this increase in transverse cardiac diameter. The independent effect of a longitudinal weight gain of 5 kg or more in accentuating the longitudinal increment in heart size could be mediated through ventricular dilatation and/or hypertrophy.

Consistent with previous cross-sectional studies of elderly persons (Cowan, 1959; Edge et al., 1964), as well as the longitudinal study of Lauder and Milne (1976), we found a small decrease in the transverse thoracic diameter in our elderly subgroup during the observation period. In our young and middle-aged men, however, transverse thoracic diameter increased over time. Longitudinal changes in AP chest diameter behaved similarly. These contrasting age-related patterns of change in chest dimensions appeared to be independent of the differing longitudinal patterns of body weight change in men of different ages. The CTR paralleled the change in heart size over the 17-year follow-up period, increasing from .407 to .426. The fact that CTR increased over five times as fast in men gaining more than 5 kg than in those who lost a similar amount suggests that longitudinal changes in this variable, like heart size, are also significantly influenced by changes in body weight. As with heart size, the aging-induced deepening of AP chest diameter could play a minor role in the longitudinal increase of CTR.

Only 3% of this clinically healthy sample developed a CTR exceeding .50 (with none larger than .51). This agrees very closely with Felson's (1973) figure of 2.1% in a younger sample and strongly supports the specificity of a CTR of .50, even in elderly individuals. Stated in another way, a ratio significantly larger than .50 should not be accepted as normal in any adult man regardless of age.

The most consistent longitudinal cardiovascular finding was the enlargement of the aortic knob, which occurred in 79% of men over the study period. This verifies, for the first time longitudinally, the results of prior cross-sectional studies (Felson, 1973; Gerstenblith et al., 1977; Henry et al., 1980) as well as our own cross-sectional results. Our .39 cm mean increase in aortic knob diameter with time did not differ significantly among the three age groups and is considerably larger than the .09 cm enlargement of the aortic root demonstrated on echocardiography in normal men between approximately ages 55 and 75 (Gerstenblith, 1977). Changes in aortic shape or rotation with age, measurement of aortic root rather than aortic knob, as well as the obvious biases inherent in any cross-sectional study may account for this difference. Again, magnification of the aortic knob due to deepening AP chest diameter over time would account for only a small part (4.8%) of the .4 cm longitudinal enlargement of this structure. Although Felson (1973) found no participant with an aortic knob larger than 4 cm, all of his participants were under age 55, and one-third were women. Otherwise, he found age-related increases in aortic dimension quite consistent with our own. Increasing aortic stiffness and changes in the peripheral vascular bed that result in increased total vascular resistance are probably etiologic factors for this age-related aortic dilatation. In contrast with heart and chest dimensions, aortic diameter appears to be influenced only minimally by alterations in body weight.

The 7% prevalence of aortic knob calcification on initial x-ray is consistent with Felson's (1973) 2.5% figure in persons under age 40. The fact that the prevalence doubled over the 17 year follow-up period, as well as the 10-year cross-sectional age difference between men with such calcification and those without it, are strong evidence that such changes can result from the "normal" aging process. Aortic tortuosity tripled in frequency between films and was present in more than half of the men by final radiologic examination.

Fraser and colleagues (1979) consider hyperinflation, increased lung markings, and diminished peripheral pulmonary vascularity as the hallmarks for the radiographic diagnosis of COPD. Using these criteria, we considered 21% of our patients to have some radiographic degree of COPD by the end of the study. In addition, most of the specific pulmonary abnormalities investigated increased in prevalence one- to four-fold over the 17-year follow-up period, suggesting that some structural changes were taking place in the lungs of these apparently healthy men over time despite the absence of clinical or spirometric abnormalities.

Hyperinflation was the most common abnormality, appearing in roughly one-quarter of the sample on first visit but increasing to over 50% on final x-ray. The presence of increased lung markings was nearly as frequent; Kerley B lines and enlargement of the pulmonary arteries occurred only rarely initially and, despite three- to five-fold increases in prevalence over time, were uncommon also on follow-up examination.

Some authors have suggested that the radiographic criteria for COPD are more sensitive than are the clinical manifestations (Sutinen et al., 1965). It has been estimated that a 20- to 30-year history of disease is required before ventilatory impairment becomes severe enough to produce dyspnea (Fraser et al., 1979; Thurlbeck, 1980). Thurlbeck (1980) observed that if one looks in minute detail, 100% of autopsies would demonstrate at least minimal degrees of emphysema. Clinically, it would appear more relevant to compare x-ray findings with pulmonary function testing than with autopsy data. Whereas none of the men in our study had even the mildest laboratory evidence of COPD by conventional spirometry, we believe that it is probably inappropriate to rely upon the usual x-ray criteria for diagnosing COPD when screening basically healthy samples such as ours. Indeed, in attempting to screen such a population for any disorder, one must be mindful of Bayes' theorem, which predicts that the diagnostic usefulness of a test for a given disease will be directly proportional to the pretest likelihood of that disease in the population tested (Rifkin & Hood, 1977). Whether the 21% with radiographic "evidence" of COPD, as well as the additional 42% whose final film demonstrated at least one pulmonary "abnormality," will develop clinical, spirometric, or autopsy evidence of COPD with greater frequency than those men with normal films remains to be determined.

In recent years more specific criteria have been suggested for the diagnosis of COPD. Using the level of the dome of the right hemidiaphragm, Nicklaus et al. (1966) claimed 94% accuracy in diagnosing severe emphysema. In our series the dome of the right hemidiaphragm was below the seventh anterior rib in only two men initially and three on the final film, supporting the high specificity of this measurement. In a clinopathologic study involving 696 patients, Thurlbeck and Simon (1978) also found the level of the hemidiaphragm to be a reliable radiographic criterion for diagnosing COPD. They further observed that the decrease in arterial vascularity corresponded with the amount of pulmonary hypertension present in the patients with severe emphysema. Ten percent of the men in our study showed minimal evidence of such diminished vascularity. The complete absence of right ventricular hypertrophy and bullae in our sample suggests that neither of these x-ray findings should ever be dismissed as normal variants.

Based on the 21% of the men in our series who were known smokers, we found no definite correlation between chest x-ray results during the study and smoking status, a finding not unexpected considering that all men with clinical or spirometric evidence of COPD were excluded from the study. The absence of any such correlation, however, further supports the nonspecificity of the radiographic findings in our sample; smokers have been found to show subtle abnormalities of pulmonary function that may not appear on conventional spirometry.

In summary, the majority of normal men will demonstrate minimal radiographic increase in heart size with advancing age, whether evaluated by transverse cardiac diameter or cardiothoracic ratio; such increases are largest in men gaining appreciable weight over time. A ratio exceeding .50 should continue to be viewed with suspicion, regardless of age. Aortic tortuosity and increase in aortic diameter occur in the majority of normal men if followed long enough, presumably due to the higher vascular resistance seen with advancing age. Radiographic suspicion of COPD occurs in about 20% of men, despite the absence of clinical or spirometric abnormalities. The imperfect specificity of these conventional x-ray criteria for COPD dictates that they be applied cautiously to populations with low prevalence of clinical lung disease.

REFERENCES

Cardiopulmonary Council of the American Heart Association. Manual for evaluation of lung function by spirometry. *Circulation*, 1982, *65*, 644A-651A.

Cowan, N. R. The heart lung coefficient in older people. *British Heart Journal*, 1959, *21*, 238-242.

Edge, J. R., Millard, F. J. C., Reid, L., & Simon, G. The radiographic appearance of the chest in persons of advanced age. *British Journal of Radiology*, 1964, *37*, 769–774.

Felson, B. *Chest roentgenology*. W. B. Saunders Company, Philadelphia, PA, 1973.

Fraser, R., Pare, G., & Peter, J. A. *Diagnosis of disease of the chest* (2nd ed., Vol. 3). Saunders and Company, Philadelphia, PA, 1979.

Gerstenblith, G., Frederiksen, J., Yin, F. C. P., Fortuin, S. N. J., Lakatta, E. G., & Weisfeldt, M. Echocardiographic assessment of a normal adult aging population. *Circulation*, 1977, *56*, 273–278.

Henry, W. H., Gardin, J. M., & Ware, J. H. Echocardiographic measurements in normal subjects from infancy to old age. *Circulation*, 1980, *62*, 1054–1060.

Krumpelman, J. L., & Burki, K. Correlation of pulmonary function with chest x-ray. *American Review of Respiratory Disease*, 1980, *121*, 217–223.

Lauder, I. J., & Milne, J. S. Longitudinal study of heart size in older people. *British Heart Journal*, 1976, *38*, 1285–1290.

Lusted, L. B., & Keats, T. E. *Atlas of roentgenographic measurements*. Year Book Medical Publishers, Chicago, IL, 1967.

Nicklaus, T. M., Stowell, D. W., Christiansen, W. R., & Renzetti, A. D., Jr. The accuracy of the roentgenologic diagnosis of chronic pulmonary emphysema. *American Review of Respiratory Disease*, 1966, *93*, 889–899.

Reid, L., & Millard, F. J. C. Correlation between radiological diagnosis and structural lung changes in emphysema. *Clinical Radiology*, 1964, *15*, 307–311.

Rifkin, R. D., & Hood, W. B., Jr. Bayesian analysis of electrocardiographic exercise stress testing. *New England Journal of Medicine*, 1977, *297*, 681–686.

Stone, J. L., & Norris, A. H. Activities and attitudes of participants in the Baltimore Longitudinal Study. *Journal of Gerontology*, 1966, *211*, 575–580.

Sutinen, S., Christoforidis, A., Klugh, G. A., & Pratt, P. C. Roentgenologic criteria for the recognition of nonsymptomatic pulmonary emphysema. Correlation between roentgenologic findings and pulmonary pathology. *American Review of Respiratory Disease*, 1965, *91*, 69–76.

Thurlbeck, W. M. Uses and limitations of the chest roentgenogram in the diagnosis of emphysema. *Practical Cardiology*, 1980, *69*, 51–58.

Thurlbeck, W. M., & Simon, G. Radiologic appearance of emphysema. *American Journal of Roentgenology*, 1978, *130*, 429–440.

J AM COLL CARDIOL
1983;1(3):887-92

Right Bundle Branch Block: Long-Term Prognosis in Apparently Healthy Men

JEROME L. FLEG, MD, FACC, DHIRENDRA N. DAS, MD, EDWARD G. LAKATTA, MD

Baltimore, Maryland

The long-term cardiac prognosis of 24 clinically healthy men with complete right bundle branch block, identified from the 1,142 men constituting the population of the Baltimore Longitudinal Study on Aging, was assessed over a follow-up period averaging 8.4 years. When compared with a control group matched for age at which right bundle branch block appeared (mean ± standard deviation 64.0 ± 13.5 years), men with right bundle branch block showed no difference in the prevalence of antecedent coronary risk factors or obstructive lung disease. The incidence of angina pectoris, myocardial infarction, valvular heart disease, cardiomegaly, congestive heart failure, advanced heart block or cardiac death in these men did not differ from that of the control group over the observation period. Furthermore, at the latest follow-up study, maximal aerobic exercise tolerance and chronotropic response to maximal exercise were not impaired in men with right bundle branch block relative to control men (9.1 ± 2.2 versus 7.3 ± 3.0 minutes and 150.3 ± 23.5 versus 147.7 ± 20.7 beats/minute, respectively). However, axis deviation leftward of −30° was present in 46% of men with right bundle branch block but in only 15% of control subjects at latest follow-up (probability [p] < 0.01). Although the PR interval lengthened by 40 ms or more developed in only 6% of control subjects over the observation period, such prolongation occurred in 29% of men with right bundle branch block ($p < 0.05$). These results support the concept that right bundle branch block in these asymptomatic men is a manifestation of a primary abnormality of the cardiac conduction system but has no demonstrable adverse effect on long-term cardiac morbidity or mortality.

Since its electrocardiographic description more than 70 years ago (1), right bundle branch block has been the subject of numerous epidemiologic investigations. In many of these early studies (2–5), the subjects were derived from hospital-based populations with heart disease; consequently, the long-term cardiovascular morbidity and mortality rates of these patients with right bundle branch block were very high. It was eventually recognized, however, that right bundle branch block by no means constituted a homogeneous clinical disorder and that its prognosis depended on the nature and extent of underlying heart disease (6–8).

Given the increased use of the electrocardiogram as a screening tool in the general population, a question of practical concern is whether the presence of right bundle branch block in apparently healthy subjects increases their likelihood for subsequent cardiac events. Conflicting answers to this question are provided by epidemiologic studies in military (9,10) and community-based (11–13) groups. Investigations in military populations with right bundle branch block generally yielded low rates of coronary disease (9,10), probably reflecting the young age and highly selected nature of their subjects. In community-based studies, right bundle branch block was generally associated with a greater frequency of organic heart disease than in the military studies. Among the community studies, only the Framingham investigators (13) employed an age-matched control population; they found a 2.5-fold increase in coronary disease and nearly 4-fold increase in congestive heart failure in 53 men and women with right bundle branch block and no initial evidence of coronary disease over a 6-year follow-up period.

This study describes the long-term cardiac prognosis of apparently healthy men with right bundle branch block, either preexisting or newly developed, identified from the Baltimore Longitudinal Study on Aging, a community-

---

From the Cardiovascular Section, Gerontology Research Center, National Institute on Aging, National Institutes of Health, Baltimore, Maryland 21224. Manuscript received August 5, 1982; revised manuscript received October 26, 1982, accepted October 29, 1982.

Address for reprints: Jerome L. Fleg, MD, Cardiovascular Section, Gerontology Research Center, National Institute on Aging, 4940 Eastern Avenue, Baltimore, Maryland 21224.

dwelling population of 1,142 men (14). In addition to comparing the incidence of subsequent cardiac events with that in an age-matched control population, we have addressed several questions to further define the natural history and pathophysiology of this conduction disorder in subjects without other evidence of cardiac disease. Is the incidence of latent coronary artery disease greater in these subjects than in control men? Do these subjects demonstrate abnormal sinus node or atrioventricular node function, or disease in another fascicle suggesting a diffuse conduction system degeneration? Is their maximal aerobic capacity or chronotropic response to exercise lower than that of their age-matched normal cohorts?

Methods

Study patients. Since 1958, the Baltimore Longitudinal Study on Aging has enrolled some 1,142 men, who have been followed up biennially (annually if 70 years or older) with extensive noninvasive testing that includes a complete history and physical examination, chest X-ray films, 12 lead rest electrocardiogram, exercise stress testing and pulmonary function tests (14). During this 23 year period, 39 men with right bundle branch block on the rest electrocardiogram were identified. Of this group, 24 men, on initial presentation with right bundle branch block, had no evidence of associated cardiac disease, as defined by angina, myocardial infarction by history or electrocardiogram, cardiomegaly, valvular heart disease or congestive heart failure, and returned for at least one subsequent visit. These 24 men constitute the subject of this report.

Definitions. *Right bundle branch block* was defined by the Minnesota Code criteria (7:2) of a limb lead QRS duration of 0.12 second or more with an R′ greater than R or an R peak duration of 0.06 second or more in either lead V_1 or V_2 (15). In men who developed right bundle branch block while under observation, the first visit was defined as the visit in which right bundle branch block was initially manifested. At each visit the presence of the following was ascertained:

Hypertension: blood pressure of 160/95 mm Hg or greater or an interim history of such an elevated pressure currently under treatment.

Hypercholesterolemia: a serum cholesterol level of more than 275 mg/100 ml.

Smoking history: a smoker was defined as a person who smoked 10 or more cigarettes per day for at least 5 years and was smoking at the onset of the study.

Diabetes: symptomatic fasting hyperglycemia requiring insulin or an oral hypoglycemic agent. Men with asymptomatic elevation of blood glucose outside standard deviations of the mean age-adjusted standards were not included in this definition.

Angina pectoris: ischemic chest pain fulfilling standard clinical criteria as evaluated by a staff cardiologist.

Myocardial infarction: a convincing history of infarction, usually verified by hospital records, or the presence of diagnostic Q waves (Minnesota Code 1:1 or 2:1),or both (15).

Valvular heart disease: valvular stenosis or insufficiency as evaluated by a staff cardiologist.

Cardiomegaly: a cardiothoracic ratio greater than 0.50 on a standard posteroanterior chest X-ray film.

Congestive heart failure: a convincing history of dyspnea, orthopnea or systemic venous congestion resolving with diuretic agents or digitalis therapy, or both, or the presence of similar clinical findings associated with radiographic evidence of pulmonary venous engorgement.

Chronic obstructive pulmonary disease: symptoms of chronic bronchitis or emphysema or physical, spirometric or radiographic signs of airway obstruction.

Sinus bradycardia: a heart rate of less than 60 beats/min.

First degree heart block: a PR interval of 0.22 second or more (Minnesota Code 6:3) (15).

Left axis deviation: an axis of the initial 0.08 second of the QRS complex of −30° or less (Minnesota Code 2:1) (15).

Atrial fibrillation: irregular atrial activity faster than 350/min with an irregular ventricular response.

Positive exercise thallium scintigraphy (performed in a subset of study subjects and control subjects): a perfusion defect appearing during maximal treadmill exercise and improving with redistribution.

Positive electrocardiographic stress test: 1.0 mm or greater J point depression with flat or downsloping ST segment for 0.08 second after the J point (Minnesota Code 11:1) (15) on double Master two-step or treadmill exercise testing, present in inferior or anterolateral leads. The validity of the exercise electrocardiogram in right bundle branch block has been previously verified for these leads (16,17).

Cardiovascular death: any death in which the underlying cause as determined by autopsy or from the death certificate was disease of the heart or blood vessels.

In order to identify more subtle longitudinal cardiovascular differences between men with right bundle branch block and control subjects than the end points just listed, we also examined the following variables in both groups:

Blood pressure: determined from the mean of four readings for systolic and diastolic blood pressures taken on each visit and averaged over the entire number of visits for each man. The slopes of systolic and diastolic blood pressure changes over time were also calculated.

Heart size: determined by the cardiothoracic ratio on the most recent chest X-ray film.

Aerobic capacity: determined on the most recent visit by maximal treadmill exercise duration in minutes, utilizing a modified Balke protocol.

Maximal exercise heart rate: determined by the maximal heart rate attained during treadmill exercise on the most recent visit.

Rest heart rate, PR interval and QRS axis: determined on the initial and most recent visits from the standard electrocardiogram. Annualized rates of change were then calculated for each of these electrocardiographic variables.

Control group. To determine whether these abnormalities occurred more frequently in asymptomatic subjects with right bundle branch block than in normal subjects, we derived a control group by age-matching each subject with right bundle branch block (using the age at initial presentation with right bundle branch block) with the two normal men whose history numbers were closest to that of the index case. Normality was defined by the absence of cardiac disease, as outlined for right bundle branch block subjects, and by the absence of any intraventricular conduction delay on electrocardiogram. All control subjects were successfully matched to within 2 years of their respective right bundle branch block index cases.

Statistical methods. The group means for continuous variables were compared for right bundle branch block and control subjects using the unpaired or paired *t* test as appropriate; discrete variables were compared by Fisher's exact test. The slope of longitudinal blood pressure change in a given subject was calculated by least squares linear regression analysis. A probability (p) value of ≤ 0.05 was considered significant for all analyses. Data are presented as mean values ± standard deviation.

Results

Age at onset of right bundle branch block. Thirty-nine men (3.4%) of 1,142 were found to have complete right bundle branch block. Eleven of the 39 had preexisting cardiac disease and another 4 were lost to follow-up after their initial visit, leaving 24 men with complete right bundle branch block without evidence of heart disease for whom follow-up information was available. Seven of these men developed right bundle branch block during the course of the Baltimore Longitudinal Study on Aging and the remaining 17 presented with this conduction disturbance. Their mean age on presentation with, or development of, right bundle branch block was 64.0 ± 13.5 years (range 33.6 to 90.5) (Table 1). The mean age of the seven men who developed right bundle branch block while under observation was 61.1 ± 8.0 years and did not differ from that of men with preexisting right bundle branch block. The mean age of the 48 control men at the onset of study was 64.2 ± 12.8 years (range 32.8 to 88.6). Follow-up duration averaged 8.4 ± 6.4 years (range 1.0 to 20.4).

Risk factors and pulmonary disease. The prevalence of antecedent coronary risk factors and obstructive pulmonary disease did not differ significantly between the two groups. Smoking was the most common risk factor in both right bundle branch block and control groups, occurring in 42 and 38% of subjects, respectively. The frequency of hypertension was 21 and 19% in the respective groups. Clinical chronic obstructive pulmonary disease was seen in 13% of men with right bundle branch block and 17% of control subjects; the mean forced expiratory volume (FEV_1) in 1 second of 76.2 ± 6.8% in subjects with right bundle branch block was also similar to the control value of 73.3 ± 9.2%. Hypercholesterolemia and diabetes were present, respectively, in 4 and 4% of men with right bundle branch block and in 12 and 4% of control subjects.

New cardiac events or abnormalities. The incidence of new cardiac events or abnormalities (angina, myocardial infarction, valvular heart disease, cardiomegaly, congestive heart failure, complete heart block or cardiac death) over the 8 year observation period was not significantly different between the right bundle branch block and control groups. Only 5 men (21%) with right bundle branch block and 10 control subjects (21%) experienced any event or abnormality. Clinical coronary disease, manifested by angina pectoris or myocardial infarction, developed in 17% of men with right bundle branch block and 13% of control subjects. Valvular heart disease, cardiomegaly and congestive heart failure occurred, respectively, in 8, 8 and 4% of men with right bundle branch block and 4, 6 and 0% of control subjects. Within the right bundle branch block group, cardiac morbidity-mortality was not related to QRS axis, QRS duration or PR interval on initial presentation with right bundle branch block group were from cardiac disease: one from an acute myocardial infarction in a 73 year old man who had developed angina during the observation period, the other from congestive heart failure complicating a sigmoid volvulus in a 78 year old man without prior heart disease. Of the 14 deaths in the control group, 4 resulted from cardiac causes, all acute myocardial infarctions, 2 of which occurred in men who did not manifest clinical heart disease while alive.

Latent coronary heart disease. The presence of latent coronary heart disease (as defined by an ischemic ST segment response to either a double Master two-step or graded treadmill exercise test) was sought in 15 (62%) patients with right bundle branch block and 43 (90%) control subjects during the observation period. (The smaller percent of subjects with right bundle branch block performing exercise reflects the belief during the early years of the study that right bundle branch block precluded accurate interpretation of the exercise electrocardiogram with the consequence that stress testing was not routinely performed.) Three subjects with right bundle branch block (20% of those exercised) and 12 control subjects (28%) were positive for ischemia by Minnesota Code 11:1. Eight men with right bundle branch block, chosen consecutively from those still active in the study, underwent thallium scanning in conjunction with maximal treadmill exercise. One man with right bundle branch block who developed coronary heart disease during follow-up and three control subjects, all of whom were asymptomatic, demonstrated perfusion defects with exercise suggestive of coronary artery disease.

Electrocardiographic abnormalities. Electrocardiographic conduction abnormalities in right bundle branch block and control subjects on both initial and most recent visits are compared in Table 1. The most striking finding

Table 1. Electrocardiographic Findings on Initial and Most Recent Visits

	Men With Right Bundle Branch Block (n = 24)			Control Group (n = 48)		
	Initial	Recent	Δ	Initial	Recent	Δ
Sinus bradycardia (%)	25	25		10	16	
Rest heart rate (beats/min)	68.0 ± 10.7	65.3* ± 8.6	−2.7 ± 9.0	71.3 ± 8.8	71.1 ± 12.2	0.1 ± 10.5
First degree AV block (%)	4	13		4	6	
PR interval (ms)	164.2 ± 25.7	190.4 ± 59.8	+26.3† ± 42.8	164.9 ± 10.0	174.6 ± 19.6	9.6‡ ± 14.0
Left axis deviation (%)	21	46§		8	15	
QRS axis (°)	2.8 ± 41.8	−9.8 ± 52.5	−12.6¶ ± 23.5	19.1 ± 31.5	11.8 ± 35.5	−7.3† ± 17.2

The percentages represent prevalence. Continuous variables are expressed as mean values ± standard deviation. Δ indicates longitudinal change in the continuous variables.
*p < 0.05 versus control; †p < 0.01 longitudinal change within group; ‡p < 0.001 longitudinal change within group; §p < 0.01 versus control; ¶p < 0.05 longitudinal change within group.
AV = atrioventricular.

was the increased frequency of left axis deviation of −30° or less in the former group on the most recent visit. In 6 of the 11 subjects with right bundle branch block and 3 of the 7 control subjects who eventually demonstrated left axis deviation, this abnormality became manifest during the observation period. Whereas no control subject displayed a QRS axis leftward of −50° on any electrocardiogram, three men with right bundle branch block showed an axis between −60° and −80° on initial presentation with right bundle branch block and two additional men developed this finding over the observation period. First degree atrioventricular block was uncommon in both groups; no subject developed a QRS axis rightward of +90°, had high degree atrioventricular block or required a pacemaker during the follow-up period.

Cardiac conduction system. In an attempt to identify subtle differences in the cardiac conduction system between right bundle branch block and control subjects, we compared heart rate at rest, PR interval and frontal plane QRS axis between groups on both initial and final visits (Table 1). Heart rate at rest was slower in the men with right bundle branch block than in control subjects on the final visit but this difference was small. PR interval was nearly identical in the two groups on the first visit and increased with age in both groups. The magnitude of this longitudinal PR interval prolongation was three times as great in right bundle branch block subjects as in control over the observation period (26.3 versus 9.6 ms, $p = 0.08$). Because 40 ms is probably the smallest increment in PR interval that can be reliably detected on a standard electrocardiogram by visual analysis, we identified the men in each group in whom such PR interval prolongation occurred over the observation period. Seven (29%) men with right bundle branch block and three (6%) control subjects developed PR interval prolongation of 40 ms or greater between initial and most recent visits ($p < 0.05$). Although the mean QRS axis of men with right bundle branch block was leftward of that of control subjects on both visits and moved further leftward at nearly twice the rate of control subjects, none of these differences was statistically significant. In both groups of men, significant leftward shifts in QRS axis were observed over time.

Blood pressure. In order to detect subtle differences in long-term blood pressure trends between the right bundle branch block and control groups, we averaged the systolic and diastolic blood pressures (four readings per visit) over the entire number of visits for each subject. Average systolic blood pressure was 131.5 ± 15.9 mm Hg in the group with right bundle branch block and 130.0 ± 16.6 mm Hg in the control group. Corresponding diastolic readings were 78.7 ± 7.0 and 78.6 ± 8.1 mm Hg. Neither reading was significantly different between the two groups. The slope of the systolic and diastolic blood pressures versus time over the 8.4 year observation period also did not differ between the right bundle branch block and control groups (0.83 ± 1.61 versus 1.21 ± 2.38 mm Hg per year for systolic blood pressure and 0.32 ± 0.89 versus −0.07 ± 1.60 mm Hg per year for diastolic blood pressure, respectively).

Heart size. The heart size of the men with right bundle branch block and control subjects on the most recent chest X-ray film was compared. Films were available on 20 of the 26 men with right bundle branch block and their 40 corresponding control subjects. Two men in each group (10 and 5%, respectively) manifested a cardiothoracic ratio of more than 0.50; the largest ratio was 0.55. Mean cardiothoracic ratio also did not differ significantly between the two groups (0.45 ± 0.05 in men with right bundle branch block versus 0.43 ± 0.04 in control subjects).

Exercise performance. Maximal treadmill exercise performance on most recent examination was assessed in clinically healthy men with right bundle branch block and control subjects with a modified Balke protocol used in our laboratory since 1975. Subjects walked at a constant speed of 3.5 mph and the incline was increased by 3% every 2 minutes, starting from the horizontal. The mean age of the 12 exercising men with right bundle branch block (69.6 ± 8.9 years) was nearly identical to that of the 20 control men (70.6 ± 9.5 years). Neither exercise duration (9.1 ± 2.2 versus 7.3 ± 3.0 minutes) nor maximal heart rate (150.3 ± 23.5 versus 147.2 ± 20.7 beats/min) differed significantly between the respective groups.

Discussion

Previous studies. Most early investigations of right bundle branch block (2–5,8) presaged a rather bleak prognosis for patients with this electrocardiographic finding primarily because the involved study groups derived from hospital wards and clinics. Even in early series (4,5,7), however, it became apparent that the outlook for a patient with right bundle branch block was by no means uniform but was strongly influenced by the patient's overall cardiac status. Reusch and Vivas (7), for example, observed a mortality rate of 32% in individuals with right bundle branch block accompanying heart disease compared with an 8% mortality rate in those with right bundle branch block and no cardiac disorder.

More recently, several epidemiologic studies have described various clinical characteristics of subjects with right bundle branch block (9–13). In a cross-sectional study of a retirement community (12), right bundle branch block with left axis deviation of less than −30° was associated with a high prevalence of cardiovascular disease but "uncomplicated" right bundle branch block was not. In a large military population (10), right bundle branch block occurred in 0.16% of apparently healthy men and was twice as common past the age of 40 years. No increase in coronary risk factors existed in these subjects. Rotman and Triebwasser (9) followed up 394 Air Force personnel with right bundle branch block for an average of 10.8 years and found that coronary heart disease developed in only 6%, reflecting the young age (mean 36 years) and highly selected character of this group.

Because neither persons in retirement communities nor highly screened military personnel can be considered representative of the general population, right bundle branch block occurring in these subjects also may not be truly representative. The community-based studies of Tecumseh and Framingham probably better approximate an unselected population. In the former investigation (11), right bundle branch block was seen primarily in elderly subjects and was associated with the appropriately high prevalence of coronary risk factors expected in this age group, but no follow-up information was provided. Data from 70 men and women with newly acquired right bundle branch block in the Framingham study (13) indicate an increased incidence of coronary disease, congestive heart failure and cardiovascular disease mortality over a mean follow-up period of 6 years compared with that in age-matched control subjects. A QRS duration of greater than 130 ms and a QRS axis left of −45° identified subgroups with a high cardiovascular risk. However, the 20 subjects who were free from associated cardiovascular abnormality at the onset of right bundle branch block had an overall favorable prognosis.

Does right bundle branch block presage subsequent cardiovascular disease and mortality? The present investigation addresses the issue whether right bundle branch block of itself increases the likelihood of subsequent cardiac events in asymptomatic subjects. In our Baltimore Longitudinal Study on Aging population, over two-thirds of all men with right bundle branch block presented without associated cardiac disease. As the number of elderly Americans continues to increase dramatically (18) and the electrocardiogram is routinely performed in large numbers of these older persons, it may be anticipated that a substantial number of clinically healthy persons with right bundle branch block will be identified. Thus, our 24 asymptomatic men with right bundle branch block, most of whom presented with this conduction abnormality on their first examination, represent a not uncommon clinical problem.

Our most pertinent finding is that cardiovascular morbidity and mortality are not increased in asymptomatic men with right bundle branch block. This reinforces the conclusions from previous studies (4–7) that it is not the right bundle branch block but the underlying heart disease that determines prognosis. The similar prevalence of coronary risk factors in men with right bundle branch block and control subjects, as well as the similar frequency with which an ischemic ST segment or abnormal thallium scintigraphic response to exercise occurred, argues against an increase in asymptomatic coronary disease in these subjects with right bundle branch block. Further evidence for the benign nature of this condition is provided by the close agreement of cardiothoracic ratios between men with right bundle branch block and control subjects at latest follow-up examination. The nearly identical blood pressures averaged over the duration of the study would appear to eliminate even preclinical hypertension as an etiologic factor for right bundle branch block in these men.

Consistent with our demonstration that cardiovascular morbidity and mortality are not increased in men with asymptomatic right bundle branch block is our finding that no impairment of aerobic exercise performance was seen over long-term follow-up. The normal heart rate response to maximal exercise suggests that chronotropic reserve is not diminished in these men despite a mild decrease in rest heart rate compared with that of age-matched control subjects.

At first glance, our cardiovascular morbidity and mortality results appear to be at odds with those of the Framingham study. However, certain salient differences in study design and patient characteristics may account for this disparity. Individuals with preexisting right bundle branch block, excluded from the Framingham series but constituting 71%

of ours, may well have a more favorable prognosis than those who develop the conduction defect later in life. Second, our series contained no women. In the Framingham study, women with right bundle branch block had twice the prevalence of cardiomegaly and congestive heart failure as did men. Finally, the 20 subjects from Framingham who were free of antecedent or coincident cardiovascular abnormalities, as were the majority of our men, had much lower cardiac morbidity and mortality rates than the other subjects, 75% remaining free from any abnormality during the follow-up period.

Etiology of associated abnormal left axis deviation. Although aging itself is associated with a leftward shift of the QRS axis (19), the high prevalence of axis deviation leftward of − 30° in our subjects with right bundle branch block cannot be explained by advanced age alone because such axis deviation was seen significantly less often in the age-matched control subjects. Because the cause of right bundle branch block in these men is not immediately obvious, an attractive hypothesis is that the right bundle branch block, as well as the associated left axis deviation, reflects isolated fibrosis of the conduction system. Pathologic studies have demonstrated a variable degree of calcification of the left side of the cardiac skeleton with advancing age (20,21). The longitudinal leftward shift of QRS axis and prolongation of PR interval in both right bundle branch block and control groups may reflect the universality of these age-related changes in the conduction system. However, the greater tendency toward left axis deviation in men with right bundle branch block, coupled with their increased incidence of PR interval prolongation and slower rest heart rate on latest follow-up electrocardiogram in comparison with the control group, lends support to an acceleration of these primary aging changes in the conduction system of these apparently healthy men with right bundle branch block.

Limitations of study. As with any epidemiologic investigation, certain limitations exist in our study. The Baltimore Longitudinal Study on Aging population evaluated here consisted primarily of upper middle class white subjects and included no women. The relatively small series studied reflects the low prevalence of right bundle branch block in the general population. Although no statistical differences existed between the right bundle branch block and control groups with respect to the development of cardiac events, it is possible that differences might appear if a large enough population was studied over a longer period of time.

Implications. We have found no long-term increase in cardiovascular morbidity or mortality in asymptomatic men with right bundle branch block when compared with age-matched control subjects. Similarly, we could detect no increase in their prevalence of coronary risk factors or latent coronary disease and no impairment of maximal aerobic performance. The increased occurrence of left axis deviation and PR interval prolongation in these men supports the concept that right bundle branch block is a manifestation of a more general abnormality of the cardiac conduction system in men without other evidence of heart disease.

We are indebted to Reubin Andres, MD, for his invaluable efforts in maintaining the Baltimore Longitudinal Study cardiovascular data files, to Sandra DePetris for technical assistance, to Elsie Beard for the statistical analysis and to Joan Griffin for secretarial assistance.

References

1. Eppinger H, Rothberger CJ. Zur Analyse Des Elektrokardiogramms. Wien Klin Wochenschr 1909;22:1091–4.
2. Graybiel A, Sprague HB. Bundle branch block: an analysis of 395 cases. Am J Med Sci 1933;185:395–401.
3. Perera GA, Levine SA, Erlanger H. Prognosis of right bundle branch block: a study of 104 cases. Br Heart J 1942;4:35–42.
4. Shreenivas, Messer AL, Johnson RP, White PD. Prognosis in bundle branch block. I. Factors influencing the survival period in right bundle branch block. Am Heart J 1950;40:891–902.
5. Messer AL, Johnson RP, Shreenivas, White PD. Prognosis in bundle branch block. III. A comparison of right and left bundle branch block with a note on the relative incidence of each. Am Heart J 1951;41:239–45.
6. Rodstein M, Gubner R, Mills JP, Lovell JF, Ungerleider HE. A mortality study in bundle branch block. Arch Intern Med 1951;87:663–8.
7. Reusch CS, Vivas JR. Clinical analysis of right bundle branch block. Am Heart J 1959;58:543–6.
8. Bauer GE. Development of bundle branch block. Am J Cardiol 1964;14:346–51.
9. Rotman M, Triebwasser JH. A clinical and follow-up study of right and left bundle branch block. Circulation 1975;51:477–84.
10. Johnson RL, Averill KH, Lamb LE. Electrocardiographic findings in 67,375 asymptomatic subjects. VI. Right bundle branch block. Am J Cardiol 1960;6:143–52.
11. Ostrander LD Jr. Bundle branch block: an epidemiologic study. Circulation 1964;30:872–81.
12. Edmunds RE. An epidemiologic assessment of bundle branch block. Circulation 1966;34:1081–7.
13. Schneider JF, Thomas HE, Kreger BE, McNamara PM, Sorlie P, Kannel WB. Newly acquired right bundle branch block: the Framingham study. Ann Intern Med 1980;92:37–44.
14. Stone JL, Norris AH. Activities and attitudes of participants in the Baltimore Longitudinal Study. J Gerontol 1966;21:575–80.
15. Rose GA, Blackburn H. Cardiovascular Survey Methods. Geneva: World Health Organization, 1968.
16. Whinnery JE, Froelicher VF Jr, Longo MR Jr, Triebwasser JH. The electrocardiographic response to maximal treadmill exercise of asymptomatic men with right bundle branch block. Chest 1977;71:335–40.
17. Tanaka T, Friedman MJ, Okada RD, Buckels LS, Marcus FI. Diagnostic value of exercise-induced ST segment depression in patients with right bundle branch block. Am J Cardiol 1978;41:670–3.
18. Annexton M. Treatable brain diseases in the elderly. JAMA 1978;240:1325–6.
19. Mihalick MJ, Fisch C. Electrocardiographic findings in the aged. Am Heart J 1974;87:117–28.
20. Lenegre J. Etiology and pathology of bilateral bundle branch block in relation to complete heart block. Prog Cardiovasc Dis 1964;6:409–44.
21. Davies M, Harris A. Pathological basis of primary heart block. Br Heart J 1969;31:219–26.

J Chron Dis Vol 35 pp 101 to 114 1982
Pergamon Press Ltd

LONGITUDINAL CHANGES IN SERUM CHOLESTEROL IN MAN: AN EPIDEMIOLOGIC SEARCH FOR AN ETIOLOGY

RICHARD J. HERSHCOPF,[1] DARIUSH ELAHI, REUBIN ANDRES, HOWARD L. BALDWIN, GARY S. RAIZES,[2] DOUGLAS D. SCHOCKEN[3] and JORDAN D. TOBIN

Clinical Physiology Branch, Gerontology Research Center, National Institute on Aging, National Institutes of Health, PHS, Department of Health and Human Services, Bethesda, and Baltimore City Hospitals, Baltimore, MD 21224, U.S.A

(*Received 10 November 1980; in revised form 5 March 1981*)

Abstract Serum cholesterol levels were determined in 1011 male participants of the Baltimore Longitudinal Study of Aging. This study presents the longitudinal changes in serum cholesterol from 1 July 1963 to 30 June 1977. Serum cholesterol values dropped 6% between 1970 and 1972. The span of this study was divided into two eras, one preceding and one following the drop. The effects of obesity, selected dietary constituents and physical activity were examined in an attempt to explain the secular change in serum cholesterol. Serum cholesterol levels were not significantly correlated to levels of weight or body mass index. Changes in weight were significantly positively correlated with changes in serum cholesterol. Overall, however, the study population did not experience a significant drop in weight and therefore, this relationship could not explain the observed drop in serum cholesterol. There were virtually no significant correlations between the absolute value of any of the dietary variables examined and the absolute level of serum cholesterol. There were significant but small changes in most dietary constituents; however, only changes in caloric intake were significantly positively correlated with changes in serum cholesterol. Because the overall change in caloric intake was small, it could explain less than 1 mg/dl of the 11 mg/dl drop. There was no overall change in physical activity. No significant correlations were found between either the level or change in physical activity and the level or change in serum cholesterol. It is concluded that neither weight nor physical activity could account for the observed changes in serum cholesterol. Changes in dietary constituents were significant and in a direction which would predict a lower serum cholesterol. However, for the group, dietary changes could not fully explain the drop in serum cholesterol. For individuals, the changes in diet poorly predicted changes in serum cholesterol. It is suggested that the observed secular drop in serum cholesterol may be due to factor(s) other than those studied.

INTRODUCTION

IN THE past decade, there has been a decline in cardiovascular mortality [1–4]. This has occurred contemporaneously with public health efforts to control alterable coronary heart disease risk factors, e.g. blood pressure, cigarette smoking, serum cholesterol and physical activity levels. There is evidence from independent cross-sectional studies that a drop in serum cholesterol has occurred in recent years [2, 5, 6]. Additionally, preliminary reports have suggested similar changes in populations followed longitudinally [7–9]. This study reports in detail the longitudinal changes in serum cholesterol levels in participants of the Baltimore Longitudinal Study of Aging (BLSA).

The BLSA has been following a large group of men for over twenty years. Serum cholesterol has been one of the parameters followed longitudinally. Height and weight were available for each visit. In addition, dietary and exercise histories were available in subgroups of this population. Therefore, not only changes in serum cholesterol but also

---

Present addresses: [1]The Rockefeller University, New York, NY 10021; [2]Scarsdale Medical Center, 259 Heathcote Road, Scarsdale, NY 16583; and [3]University of South Florida College of Medicine, Tampa, FL 33612 changes in factors which were the focus of public health measures have been monitored. This study presents the longitudinal changes in serum cholesterol over the 14-yr span, 1963–1977, and investigates the effects of obesity, diet, and physical activity in a well-characterized population.

METHODS

Subjects

From its inception in 1958 until 30 June, 1977, there have been a total of 1088 male participants in the BLSA. Of these, 1012 had visits between 1 July 1963 (when systematic serum cholesterol analyses began) and 30 June 1977. Only one participant in the latter period failed to have a serum cholesterol determination. Thus, 1011 participants had a total of 5127 cholesterol determinations over a span of 14 yr. The participants were all self-recruited, community-dwelling volunteers, ranging in age from 17 to 102 yr. These men spent 2½ days in the Gerontology Research Center (GRC), returned at 12 to 24-month intervals, and underwent a battery of clinical, physiologic, and psychologic tests. Participants were generally of middle and upper-middle socio-economic status and involved in sedentary work. Social and demographic characteristics of this population have been described in greater detail elsewhere [10].

Serum cholesterol methodology

On each participant visit, a serum cholesterol sample was drawn after an overnight fast. The reference method was a modification of the method of Abell & Kendall [11]. From 1963 to April 1969, all samples were done by this method at the GRC. Beginning in April 1969, all samples went to BioScience Laboratories, Van Nuys, California (BS) and 1/8 of the samples were split and run in tandem at the GRC, using the same reference method that had been in use since 1963. From March 1969 to July 1970, the BS analyses were made using the method of Kessler [12]. From July 1970 to July 1977, the BS method was that of Wybenga [13]. Based on the samples run in tandem, three separate ratios were used to convert the BS values to the GRC standard: (1) March 1969 to September 1972, the BS values were divided by 0.920; (2) from October 1972 to June 1975, 0.965; (3) from July 1975 to June 1977, 0.937. The standard errors of these ratios were 0.005, 0.004 and 0.004, respectively, and the numbers of paired samples used to compute these ratios were 194, 194 and 144, respectively.

Reproducibility of the cholesterol analyses was estimated by including a sample from a common plasma pool with each analytical run. From 1962 to 1970, 22 separate commercial plasma pools (Hyland Laboratories, Los Angeles, California) were used in 301 individual assays. The mean coefficient of variation (CV) of these runs averaged 2.6%. From 1970 to 1977, three large plasma pools prepared by us were used for analyses by our laboratory and by BS. The mean CV in 96 samples in our laboratory was 4.3% and in 107 samples by BS was 3.8%.

Since the cholesterol analyses were conducted over a 14-yr period, we examined the data by three techniques for the possibility of methodologic drift or for blocks of time during which technical errors may have occurred. The techniques were: (1) A random sample of frozen and of lyophilized plasma samples ($N = 385$) over the entire time period were reanalyzed at the end of the study and compared to the original analyses. (2) The mean age-specific cholesterol values over the time span were examined. (3) Advantage was taken of the longitudinal nature of the data to examine for systematic deviation of the data at any one period of time. The technique was to compute the regression of cholesterol on age for each subject. The deviation of each datum from the value predicted from the regression was computed. These deviations from all the subjects were then grouped by time period. Thus, if in any block of time, cholesterol values were being measured incorrectly (say, too low), then the mean deviation for that time period would differ significantly from zero.

Using those three methods, we examined the 28 six-month time intervals in the study. Only one of the 28 time periods, the first six months of 1966, showed consistent deviations by all three methods.

1. The frozen and lyophilized samples on repeat analysis averaged 4.5% less than the original analyses, but this was, with the exception of the first half of 1966, quite consistent. The ratio of the repeat analysis to the original analysis for the 22 samples during the first half of 1966 averaged 0.88; this was about 7% lower than the mean ratio of the other time periods.
2. The mean age-specific cholesterol values for 154 subjects tested during the first six months of 1966 was the lowest value of any of the six-month periods during the 1963-1970 time interval, the time interval during which cholesterol values were relatively high (see Results section). The mean value for that six-month period was 214 mg/dl, while the other six-month intervals immediately preceding and following that time averaged 234 mg/dl.
3. The cholesterol values in that six-month period averaged 10.3 mg/dl below the predicted value computed by individual longitudinal slope analysis. This was a larger deviation than that of any of the other six-month periods.

To summarize, by all three methods, the first half of 1966 was the only time period which had consistent and large deviations. The explanation for this problem is not known but the error was so large that the decision was made to omit this period from the data presented. With this six-month interval eliminated, the population sample drops from 1011 to 1001 participants with 4973 serum cholesterol determinations.

Cortisone glucose tolerance test

During the 14-yr span, participants periodically were given this test as part of their multiphasic examination. All plasma cholesterol determinations done when the participants were cortisone-primed have been eliminated, leaving 4334 determinations on 993 participants.

Clinical classification

In order to define changes in serum cholesterol which are not secondary to specific diseases and medications, the following were used as the basis for excluding data from the further analysis:

1. Coronary heart disease: definite evidence of coronary heart disease [14] on any visit excluded all data on that participant.
2. Diabetes mellitus: overt diabetes as evidenced by a history of antidiabetic medication at any time or by two fasting plasma glucose values ≥ 140 mg/dl excluded all participant data.
3. Anti-lipid medication: history of anti-lipid medication excluded all participant data.
4. Thyroid medications: history of thyroid medication at a visit excluded data for that visit.
5. Hyperthyroidism or hypothyroidism: clinically evident disease at a visit excluded data for that visit.
6. Systemic corticosteroids: use of systemic corticosteroids at a visit excluded data for that visit.
7. Systemic sex steroids: use of systemic sex steroids at a visit excluded data for that visit.
8. Kidney disease: defined as a creatinine clearance less than the first percentile for age [14] confirmed by a serum creatinine worse than the first percentile for age. Data were excluded for only those visits which met the criteria.
9. Liver disease: biochemically confirmed liver disease at a visit excluded data for that visit.

10. Miscellaneous: data for a visit was excluded if a participant

(a) was not fasted,
(b) had history of recent surgery or illness, or
(c) had an established diagnosis of malignancy with evidence of systemic effect, i.e. weight loss, anemia.

Visit and participant exclusions by cause are summarized in Table 1.

In this way, a normal group of 783 participants with 3088 serum cholesterol determination was defined. That this normal group differs somewhat from the total group (Table 2) is not surprising; however, this difference is not very great.

Obesity

Height and weight were measured with subjects wearing a standard cotton hospital gown, without shoes or slippers. As a measure of obesity, weight was corrected for height, using the body mass index (BMI) defined as weight (kg) divided by the square of the height (m).

Dietary intake

Dietary intake was assessed using a seven-day dietary diary. Nutritionists instructed the subjects in the keeping of the diary, used food models to show them how to estimate portion size, and accompanied them at lunch during their stay. The subjects were instructed to record everything taken into their mouths and swallowed during their first 'normal' week after returning home. The completed record was mailed back, reviewed, corrected if necessary and coded. Conversion of food intake to dietary constituents was based on the food lists in the U.S. Department of Agriculture Handbook 8. The method and dietary characteristics of this population have been presented in detail elsewhere [15, 16].

Physical activity

Estimates of physical activity were derived from an interview or questionnaire covering specific activities at home, at work, at recreation, and variations in activity patterns such as trips, seasonal sports and the like. Total daily energy expenditures were calculated for each subject by use of predetermined values for each activity as previously described [15].

TABLE 1. NUMBER OF VISITS AND PARTICIPANTS EXCLUDED FROM ANALYSIS BY CAUSE

	Number of visits	Number of participants
Coronary heart disease	934	154
Diabetes mellitus	277	42
Antilipid medication	76	13
Thyroid medication	210	53
Hyper- and hypothyroidism	11	10
Systemic corticosteroids	60	28
Systemic sex steroids	51	23
Kidney disease	41	17
Liver disease	16	5
Miscellaneous	26	22
Total*	1246	210

*Since some subjects had more than one reason for exclusion, the sum of the individual exclusions is greater than the total number actually excluded.

TABLE 2. EFFECT OF AGE ON SERUM CHOLESTEROL LEVELS IN THE TOTAL POPULATION AND IN THE NORMAL SUBGROUP (SEE TEXT)

Age group (yr)	Serum Cholesterol* (mg/dl) Total group N	Mean	SEM	Normal group N	Mean	SEM
20–39	249	206	2.1	245	206	2.1
40–59	373	236	1.8	306	234	1.9
60–79	319	229	2.0	204	225	2.6
80–102	50	220	4.7	26	214	6.6
17–102†	993	225	1.2	783	222	1.3

*·˙ Each participant is represented only once by his mean cholesterol level and is grouped by his mean age.

†Includes 2 participants less than 20 yr of age.

Statistical methods

A participant was represented only once within any time period presented. For certain analyses, the one value for that participant was the mean value of all the determinations within the period under observation. In those instances the variables are denoted by a '˙'. In the longitudinal analyses of changes in serum cholesterol (Fig. 4), the annual rate of change for serum cholesterol was computed as the slope of the regression line for each subject with three or more 'normal' data points. In the longitudinal analysis of secular change each participant was represented by the mean of available determinations within each time period. The differences were assessed by paired *t*-test. Simple linear correlations and regressions were used to assess the relationship between variables for those same individuals.

RESULTS

Cross-sectional age differences, 1971–1977

The distribution of selected centiles of serum cholesterol by age over the time interval 1971–1977 (ERA 2, see below) is shown in Fig. 1. Higher serum cholesterol values are

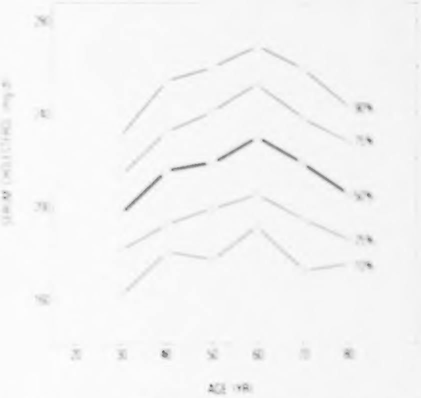


FIG. 1. Selected centile distribution of serum cholesterol for each age decade. These values represent the time period 1971–77 (ERA 2), and the number of subjects for each age group is given in Table 3.

found with increasing age from 25 to 64 yr and then lower values from 65 to 84 yr. Concentrations are symmetrically distributed within each age group and the variance does not differ with age.

Temporal differences

When examined as successive cross-sectional studies, serum cholesterol values were fairly constant from the periods 1964 to 1970. Between the 1970 and 1972 periods, there was a 6% drop in serum cholesterol and since then there has been little change (Table 3). This is evident in all three broad age categories (Fig. 2). Because of this drop in serum cholesterol, the total span of the study has been separated into two eras: 1 July 1963 to 30 June 1971 (ERA 1) and 1 July 1971 to 30 June 1977 (ERA 2). Cross-sectional differences in serum cholesterol between ERA 1 and ERA 2 were found in all age decades studied (Fig. 3).

Longitudinal changes

The longitudinal analysis of serum cholesterol within ERA 1 shows that changes in cholesterol levels within individuals grouped by age follow the pattern suggested by the cross-sectional analysis, that is, increasing cholesterol values in the younger adult years and decreasing values in the later years (Fig. 4A).

TABLE 3. SUCCESSIVE CROSS-SECTIONAL AGE DIFFERENCES IN SERUM CHOLESTEROL

Age (yr)		1964	1966	1968	1970	1972	1974	1976	ERA 1 1963–1971	ERA 2 1971–1977
17–24	X								188	180
	SEM								11.2	9.7
	N								12	11
25–34	X	216	214	230	213	203	203	195	217	198
	SEM	4.9	8.9	7.7	4.6	3.8	3.6	3.0	3.7	2.4
	N	20	16	23	53	57	81	97	78	123
35–44	X	229	224	232	228	218	219	212	228	217
	SEM	5.5	5.4	4.7	4.1	4.5	4.2	3.7	3.1	3.2
	N	65	59	83	66	49	66	74	119	92
45–54	X	234	236	243	243	221	221	224	240	221
	SEM	4.9	5.3	3.7	4.2	3.5	3.6	3.6	3.0	2.8
	N	76	65	114	108	101	93	86	165	124
55–64	X	238	242	236	238	228	226	228	237	229
	SEM	6.4	5.3	4.2	4.4	3.9	3.5	3.4	3.3	3.0
	N	54	50	84	80	87	98	100	116	116
65–74	X	229	230	234	226	218	218	224	232	220
	SEM	6.3	5.6	5.9	4.8	4.5	4.5	5.9	4.0	4.0
	N	37	45	54	63	55	60	50	87	71
75–84	X	220	204	222	227	206	206	217	219	209
	SEM	14.5	7.3	7.3	5.6	5.2	5.8	6.4	5.4	4.2
	N	13	20	17	29	34	38	29	38	49
85–102	X								213	206
	SEM								17.5	9.1
	N								6	8

The two-year period labeled '1964' began 1 July 1963, and ended 30 June 1965; other two-year periods are similarly labeled. ERA 1 began 1 July 1963, and ended 30 June 1971; ERA 2 began 1 July 1971, and ended 30 June 1977. Each subject appears only once in each period.

For ERA 1 and ERA 2 the mean cholesterol value for each subject during the appropriate interval of time was used.

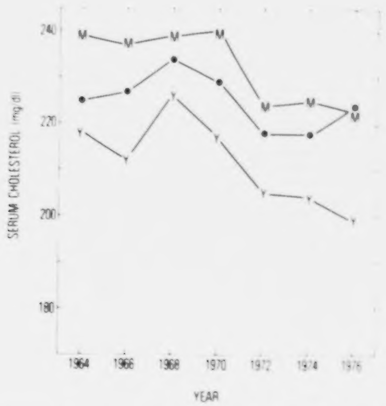


FIG. 2. Temporal differences in mean serum cholesterol levels in three age groups. Y = young group (20–39 yr), M = middle-aged group (40–59 yr), O = older group (60–79 yr).

Longitudinal analysis which encompasses the period of the cross-sectional drop in cholesterol, that is, 1969–1977, shows that the pattern of cross-sectional differences (Fig. 3) are strikingly evident longitudinally as well (Fig. 4B).

Secular change

To reiterate, both a cross-sectional drop and a longitudinal change in the serum cholesterol values of the participants of the BLSA occurred between 1970 and 1972. Longitudinal changes can be the result of three different factors: physiologic aging, methodologic drift, or environmental change. By epidemiologic convention, change in a variable secondary to environmental influences is termed secular.

In this instance, physiologic aging can be discounted because of the rapidity of the drop, the decrease in all age groups, even the younger ones, and the lack of *projected contiguity*. Projected contiguity can be conceptualized as the expectation that the final value for a group of individuals as they age, say from 30 to 40 yr of age, should approximate the initial value for a group which has aged from 40 to 50 years of age. The lack of projected contiguity for all age groups in the period encompassing the drop in serum cholesterol values is readily apparent (Fig. 4B).

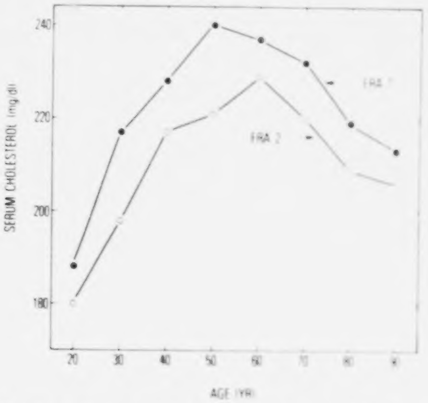


FIG. 3. Temporal drop in mean serum cholesterol values in subjects aged 17–102 yr.

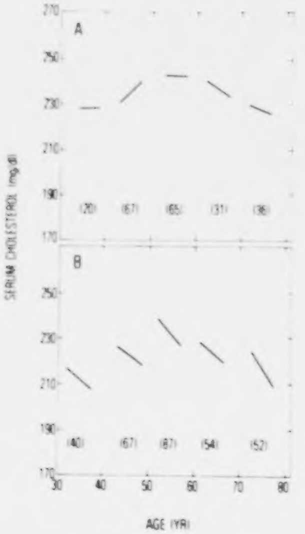


FIG. 4. Longitudinal changes in serum cholesterol concentration. Longitudinal results are represented by line segments which indicate the mean slope of changes in serum cholesterol for each age decade. Each line is drawn with the midpoint at the mean cholesterol, with the length along the abscissa representing the mean time span over which the longitudinal data were collected. The upper panel (A) presents the longitudinal change during the period prior to the drop in cholesterol (1963–1971). The lower panel (B) presents the longitudinal change which occurred during the period in which cholesterol levels fell (1969–1977). Number of subjects used to compute each mean slope is given in parentheses.

Utilizing the technique of reanalysis of stored frozen and lyophilized samples (see Methods), we could identify no methodologic shift to explain the drop in serum cholesterol. By elimination, it is likely that some change in the environment may have been responsible for the drop in serum cholesterol values.

In an attempt to explain this secular change, potential variables which could affect serum cholesterol were examined; these included (a) obesity, (b) dietary constituents, and (c) physical activity. Three analyses were applied to each variable:

1. The *absolute level* of the variable was correlated with the *absolute level* of serum cholesterol (Table 4, variable vs cholesterol). Values for the same individuals are correlated for each ERA separately. It is expected that a truly significant correlation would be consistent and therefore, be found in both ERA 1 and ERA 2.

2. The change in the variable between the ERA 1 and ERA 2 was correlated with the change in serum cholesterol (Table 4, Δ variable vs Δ cholesterol). This method tests the hypothesis that, for individuals, the change in serum cholesterol is related to a change in the variable of interest.

3. The change in the absolute value between ERA 1 and ERA 2 was computed for each individual. Even if strong relationships were found in the first two analyses, if there was no significant change in the variable of interest from ERA 1 to ERA 2, it could not explain the drop in serum cholesterol.

The subjects were grouped by mean age in ERA 1 (age). Only those subjects who had contemporaneous dietary histories and serum cholesterol determinations in both ERA 1 and ERA 2 were used in the obesity and dietary analyses. Serum cholesterol values for this group are tabulated with these variables (Table 4). A similarly constructed subgroup was formed to examine the effect of physical activity.

(a) *Obesity* (Table 4). For both ERA 1 and ERA 2, there are essentially no signficant correlations between the absolute levels of serum cholesterol and the level of either weight or BMI. In all age groups except the very old (VO), there was a significant positive correlation between a change in serum cholesterol and a change in weight or BMI. On the average, a change of 1 kg in weight resulted in a 2 mg/dl change in serum cholesterol. Although significant, the correlation explains less than 15% of the variance in changes in serum cholesterol. Between ERA 1 and ERA 2, the groups of subjects below age 60 did not have any significant overall change in obesity, but the groups over 60 had a small but significant mean weight loss. Indeed, some patients gained weight (max = 14 kg) while others lost weight (max = 16 kg). However, because there was little or no change in obesity overall, the secular drop in serum cholesterol in this population cannot be explained by a change in obesity.

(b) *Dietary constituents* (Table 4). For both ERA 1 and ERA 2, there were virtually no significant correlations between the absolute values of any of the dietary variables and the absolute level of serum cholesterol. Total calories was the only variable in which there was a significant correlation between a change in the variable and a change in serum cholesterol. The correlation was significant in the group as a whole and in the middle age group. However, this correlation explains less than 5% of the variance in the change in cholesterol. Between ERA 1 and ERA 2, there were significant but small decreases in the intake of total calories, saturated fatty acids (SFA), per cent of total calories derived from SFA, dietary cholesterol, and dietary cholesterol/1000 calories. Over the same period, there were significant but small increases in the intake of polyunsaturated fatty acids (PUFA), in the per cent of total calories derived from PUFA, and in the ratio of polyunsaturated to saturated fatty acids (P/S). Although there were significant changes in these dietary constituents, only change in total calories was significantly related to change in serum cholesterol. The magnitude of the secular change in total calories is so small that it can account for a drop of less than 1 mg/dl in serum cholesterol.

(c) *Physical activity* (Table 4). The subjects' absolute energy expenditure computed as calories/day was not consistently correlated with the absolute level of serum cholesterol. Only in the older group were changes in caloric expenditure correlated with changes in serum cholesterol. As there was no significant change in overall physical activity calories between ERA 1 and ERA 2, this variable cannot be responsible for the observed drop in serum cholesterol.

DISCUSSION

Serum cholesterol has commanded a great deal of attention as an alterable coronary risk factor. Most studies of serum cholesterol and age have been cross-sectional. This longitudinal study presents data demonstrating that changes in serum cholesterol with age coincide with the cross-sectional differences seen in different age groups. This suggests that the lower levels in the older age groups are not solely a function of selective or differential mortality, i.e. that those individuals with the highest serum cholesterol values die leaving only those with lower values to be included in the older age groups.

In addition, the temporal differences in serum cholesterol values noted in the independent cross-sectional studies during the last decade have been documented to occur longitudinally in this study population [2, 5, 6]. However, the reason for this change remains largely unexplained.

The effect of obesity on serum cholesterol is controversial [17]. The disparate results stem from at least four separate factors. One is the use of different obesity measures. The second is the confounding covariance of age with obesity and with cholesterol. The third is the study of different populations. The fourth is the fact that the correlation between obesity and serum cholesterol levels is very weak if it exists at all. The present study fails to find any consistent significant correlation between obesity and serum cholesterol level.

Reports have been much more consistent in relating changes in weight (either gain or loss) to changes in serum cholesterol [18–20]. However, our group as a whole showed

TABLE 4 SECULAR CHANGES IN SERUM CHOLESTEROL, OBESITY, DIETARY CONSTITUENTS AND PHYSICAL ACTIVITY AMONG DIFFERENT AGE GROUPS

Variable†	Age group‡	N	ERA 1		ERA 2		ERA 2 ERA 1		Correlation coefficients		
									Variable vs cholesterol		Δ Variable vs Δ cholesterol
			Mean	SEM	Mean	SEM	Mean	SEM	ERA 1	ERA 2	
Serum cholesterol (mg dl)	Y	56	216	4.1	207	3.6	−9	4.0*			
	M	158	236	3.0	227	2.8	−9	2.1**			
	O	89	231	3.5	214	3.2	−17	2.6**			
	VO	6	210	9.5	202	8.3	−8	6.2			
	All	309	230	2.0	219	1.9	−11	1.5**			
Weight (kg)	Y	56	80.4	1.71	81.0	1.63	0.6	0.63	0.248	−0.013	0.299*
	M	158	79.3	0.73	79.3	0.74	−0.0	0.28	−0.039	−0.063	0.208**
	O	89	75.8	1.04	74.7	1.10	−1.1	0.32**	−0.138	0.019	0.369**
	VO	6	72.5	3.93	69.6	3.59	−2.9	0.88**	−0.348	−0.077	−0.010
	All	309	78.3	0.58	78.1	0.60	−0.3	0.21	−0.011	−0.010	0.278**
Body mass index (kg m^{2})	Y	56	25.1	0.40	25.3	0.38	0.20	0.19	0.282*	0.083	0.315*
	M	158	25.4	0.19	25.5	0.20	0.05	0.09	0.084	0.055	0.198*
	O	89	24.8	0.29	24.5	0.31	−0.22	0.10*	−0.136	0.004	0.367**
	VO	6	24.7	1.15	24.0	1.05	−0.76	0.25**	−0.265	−0.106	0.051
	All	309	25.2	0.15	25.1	0.15	−0.02	0.06	0.063	0.067	0.274**
Calories (per day)	Y	56	2524	80	2378	63	−146	64*	−0.001	−0.034	0.224
	M	158	2319	37	2273	37	−46	30	−0.039	−0.003	0.193*
	O	89	2221	51	2137	52	−84	34*	−0.081	−0.052	0.015
	VO	6	1864	140	1883	191	18	73	0.206	0.361	0.236
	All	309	2319	29	2245	28	−74	22**	−0.059	−0.010	0.160*
Polyunsaturated fatty acids (g day)	Y	56	15.2	0.65	15.8	0.60	0.7	0.57	−0.054	−0.106	−0.022
	M	158	14.6	0.37	16.1	0.44	1.5	0.43**	−0.040	−0.067	0.145
	O	89	14.3	0.62	14.1	0.55	−0.2	0.47	0.023	−0.128	0.071
	VO	6	10.8	1.71	12.5	2.28	1.7	1.90	0.736	0.345	0.670
	All	309	14.5	0.29	15.4	0.30	0.9	0.28**	−0.015	−0.052	0.116
Per cent of calories from polyunsaturated fatty acids	Y	56	5.6	0.17	6.1	0.17	0.54	0.17**	−0.074	−0.092	−0.222
	M	158	5.8	0.12	6.5	0.13	0.62	0.13**	−0.036	−0.068	0.004
	O	89	5.9	0.20	6.0	0.16	0.05	0.17	0.135	−0.097	0.159
	VO	6	5.3	0.62	6.3	0.97	0.98	0.67	0.609	0.318	0.764
	All	309	5.8	0.09	6.3	0.09	0.45	0.09**	0.036	−0.041	0.038

Saturated fatty acids (g day)	Y	56	46.8	1.90	41.5	1.38	−5.3	1.88**	−0.053	−0.058	0.188
	M	158	40.4	0.87	38.4	0.81	−1.9	0.70**	−0.037	−0.035	0.046
	O	89	38.4	1.34	34.8	1.23	−3.5	0.92**	−0.006	−0.093	0.076
	VO	6	36.3	4.04	30.4	3.64	−5.9	2.93	0.381	0.538	−0.184
	All	309	40.9	0.70	37.8	0.62	−3.1	0.57**	−0.060	−0.042	0.093
Per cent of calories from saturated fatty acids	Y	56	16.5	0.34	15.6	0.25	−0.85	0.39*	−0.068	−0.032	0.120
	M	158	15.6	0.20	15.3	0.21	−0.32	0.18	−0.009	−0.037	−0.150
	O	89	15.4	0.30	14.5	0.30	−0.82	0.23**	0.065	−0.077	0.094
	VO	6	17.6	1.32	14.4	0.68	−3.18	1.21*	0.288	0.730	−0.253
	All	309	15.7	0.15	15.1	0.15	−0.62	0.14**	−0.027	−0.034	−0.017
Ratio of poly-unsaturated fatty acids to saturated fatty acids (P S)	Y	56	0.34	0.013	0.39	0.013	0.050	0.0148**	−0.046	−0.088	−0.247
	M	158	0.38	0.010	0.43	0.010	0.055	0.0110**	−0.016	−0.054	0.050
	O	89	0.39	0.015	0.42	0.014	0.036	0.0142*	0.102	−0.072	0.046
	VO	6	0.30	0.042	0.42	0.061	0.117	0.0558	0.548	0.089	0.703
	All	309	0.37	0.007	0.42	0.007	0.050	0.0075**	0.050	−0.033	0.014
Dietary cholesterol (mg)	Y	56	523	25	470	22	−53	19**	−0.058	−0.046	0.206
	M	158	528	14	486	14	−42	13**	−0.062	−0.048	−0.020
	O	89	497	20	455	18	−42	13**	−0.023	0.034	0.147
	VO	6	458	47	431	79	−26	42	−0.597	0.057	0.205
	All	309	517	10	473	10	−44	8**	−0.047	−0.006	0.062
Dietary cholesterol 1000 calories	Y	56	208	8	199	8	−9	6.5	−0.078	−0.052	0.054
	M	158	228	5	216	6	−12	4.9*	−0.053	−0.055	−0.109
	O	89	227	8	217	9	−10	5.5	0.034	0.077	0.176
	VO	6	248	26	230	27	−18	16.0	−0.833**	−0.376	−0.040
	All	309	224	4	213	4	−11	3.2**	−0.018	−0.005	−0.010
Physical activity (calories)	Y	84	1233	46	1237	36	4	38	0.222*	0.122	0.135
	M	218	1080	19	1084	19	4	15	−0.108	−0.205**	0.030
	O	89	923	22	944	31	21	24	−0.079	−0.060	0.216*
	VO	3	971	192	727	62	−244	147	—	—	—
	All	394	1076	16	1082	16	6	13	−0.037	−0.079	0.092

ERA 1 = 1 July 1963 to 30 June 1971.
ERA 2 = 1 July 1971 to 30 June 1977.
Y (young) = 20–39 yr old. **M** (middle aged) = 40–59 yr old. O (older) = 60–79 yr old. VO (very old) = 80–87 yr old.
***P** < 0.05. ****P** < 0.01.
† = each participant is represented only once in each ERA by the mean of all available determinations in that ERA.
‡Participants are grouped by mean age (āge) in ERA 1.

little or no change in weight. Thus, even though change in obesity is a significant modulator of serum cholesterol levels, it cannot explain the observed drop between ERA 1 and ERA 2.

The diet-heart controversy has continued [21, 22]. Although this study to date has not looked at the end point of coronary mortality, certain conclusions can be drawn about diet in the free living population and its effect on serum cholesterol. In this study population, there were clearly significant but small changes in the diet, such that one might expect a decrease in serum cholesterol. Three separate formulae were used to estimate the expected dietary effect on serum cholesterol in the BLSA participants (Table 5) [23-25]. Thus, although diet has been demonstrated to be an effective modulator of serum cholesterol, it does not seem to be the major operative factor in the presently observed change in serum cholesterol. It is interesting to note that the present study and other epidemiologic studies in Americans have not shown dietary factors to be correlated to serum cholesterol levels [26-28]. However, cross-cultural and elegant metabolic ward studies by Keys and co-workers have confirmed the effect of diet on serum cholesterol [23, 29]. This discrepancy may be due to the relatively small range in the pertinent dietary variables in Amecicans or to a threshold effect above or below which different levels are unimportant, or to the recognized lack of precision in the quantification of dietary intake. However, all the dietary constituents considered in the present analysis covered an approximately ten-fold range (except for per cent calories from SFA and PUFA). No threshold effect was observed in the present data. Careful efforts were taken to assure the quality of dietary records. However, actual dietary changes may be greater than the dietary histories indicate. The histories may not reflect the trend toward selecting less fatty grades of meats, to trimming more fat off meat both in the market and at home, and to undetermined changes in the fatty acid composition of certain manufactured foodstuffs.

A recent publication has suggested the need to look at longitudinal changes in diet and cholesterol in order to reduce interindividual variation [24]. The present study still fails to find any significant relationship between changes in the dietary constituents measured (except for calories) and changes in serum cholesterol. It is further likely that the effect of change in caloric intake is operative via the mechanism of change in weight. There was no significant correlation between the actual changes in serum cholesterol and the changes in serum cholesterol predicted by the Keys' formula [23] within ERA 1 and ERA 2 ($r = 0.091$ and 0.016 respectively). In this study, then, the drop in serum cholesterol is not adequately explained on the basis of dietary changes. Under metabolic ward conditions, such dietary effects occur. However, in a population eating under free-living

TABLE 5. ACTUAL AND PREDICTED CHANGES IN SERUM CHOLESTEROL (ERA 2 MINUS ERA 1)

Formula	ERA 2 minus ERA 1		Correlation Coefficient
	Actual change	Predicted change	Actual vs Predicted
Keys* [23]	−11.4	−2.9	−0.023
Jacobs† [24]	−11.4	−2.7	−0.022
Hegsted‡ [25]	−11.4	−5.5	0.034

$N = 309$, all ages.

*$\Delta Chol = 1.35\,(2\Delta S - \Delta P) + 1.5\,\Delta Z$.

†$\Delta Chol = 1.26\,(2\Delta S - \Delta P) + 1.5\,\Delta Z$.

‡$\Delta Chol = 2.16\,\Delta S - 1.65\,\Delta P + 6.77\,\Delta C - 0.53$.

Chol is serum cholesterol in mg/dl; S and P are percentages of total calories per day consumed as saturated and polyunsaturated fat, respectively; C is dietary cholesterol in dg/day, and Z is the square root of dietary cholesterol in mg/1000 calories.

conditions, the 'spontaneous' mild or moderate dietary changes which occurred over the years that we studied correlate poorly with changes in serum cholesterol.

Physical activity has been proposed to be a risk factor negatively correlated with coronary mortality. Our data suggest that physical activity does not exert its purported beneficial effect via changes in serum cholesterol. The lack of relationship of physical activity to serum cholesterol level has been seen previously [19, 30]. The present study suggests that this concept can be extended to include the fact that moderate changes in physical activity are not likely to change serum cholesterol significantly. However, it is likely that an increase in physical activity large enough to induce weight loss could lead to a decrease in serum cholesterol.

The present study has demonstrated both a cross-sectional and longitudinal drop in serum cholesterol over the last decade in a select population. The change in serum cholesterol was felt to be due to environmental factors. Since changes in obesity, dietary intake and physical activity could not fully explain this change, it would seem that some as yet undetermined factor(s) might be responsible. In addition, the study supports the view that change in obesity can effect changes in serum cholesterol. However, within the limits of the epidemiologic tools at hand, the spontaneous changes in dietary intake and physical activity found in this free-living population do not seem to have a significant effect on serum cholesterol levels.

ADDENDUM

Subsequent to submission of this paper, a pertinent article has been published (Shekelle RB, Shyrock AM, Paul O *et al.*: Diet serum cholesterol and death from coronary heart disease. **N Engl J Med** 304: 65–70, 1981). As part of this study of 1900 American males aged 40–55 years, the authors examined the relationship of serum cholesterol to dietary and obesity factors at the initial visit (1957) and again one year later. The authors describe small but significant relationships in the absolute levels of SFA, dietary cholesterol and BMI with the absolute level of serum cholesterol. They found similar significant relationships between changes in these variables and the changes in serum cholesterol that occurred over the one year period of observation.

Acknowledgements The authors express their gratitude to the participants in the ongoing Longitudinal Study of Aging at the Gerontology Research Center who volunteered to be subjects in these studies. We also express our appreciation to Ms Helen Burns and Ms Dolores Wehner for excellent secretarial assistance.

REFERENCES

1. United States Department of Health, Education and Welfare, National Heart, Blood Vessel, Lung, and Blood Program: **Fourth Report of the Director of the National Heart, Lung, and Blood Institute.** DHEW Publication No. [NIH] 77-1170, Washington, DC, U.S. Government Printing Office, 1977.
2. **Proceedings of the Conference on the Decline in Coronary Heart Disease Mortality, National Heart, Lung, and Blood Institute, NIH, Bethesda, Maryland, October 24–25, 1978.** Havlik RJ, Feinleib M (Eds) U.S. Department of Health, Education and Welfare, Public Health Service, NIH Publication No. 79-1610, May, 1979.
3. Walker WJ: Coronary mortality: what is going on? **J Am Med Assoc** 227: 1045–1046, 1974.
4. Stern MP: The recent decline in ischemic heart disease mortality. **Ann Intern Med** 91: 631–640, 1979.
5. Abraham S: Total serum cholesterol levels of adults, 18–74 years, United States, 1971–1974. **Vital and Health Statistics.** Series 11 No. 205, DHEW Publication No. [PHS] 78-1652, Washington, DC, U.S. Government Printing Office, 1978.
6. Levy R: Press briefing on arteriosclerosis program goal, July 13, 1977. Information Office, National Heart, Lung, and Blood Institute, Bethesda, MD 20014, U.S.A.
7. Taylor HL, Romo M, Jacobs DR, *et al.*: Secular changes in coronary heart disease risk factors. **Circulation** 52 (Suppl II): 96, abstract, 1975.
8. Harlan WR, Oberman A, Carlson S, *et al.*: Longitudinal assessment of serum lipoproteins and cholesterol. **Clin Res** 27: 221, abstract, 1974.
9. Hershcopf, R, Baldwin H, Elahi D, *et al.*: Cholesterol and aging: from the Baltimore Longitudinal Study of Aging. **J Gerontol** 18: 81, abstract, 1978.
10. Stone JL, Norris AH: Activities and attitudes of participants in the Baltimore Longitudinal Study. **J Gerontol** 21: 575–580, 1966.
11. Abell LL, Levy BB, Brodie, BB *et al.*: A simplified method for the estimation of total cholesterol in serum and demonstration of its specificity. **J Biol Chem** 195: 357–366, 1952.

12. Kessler G: Automated techniques in lipid chemistry. **Adv Clin Chem** 10: 45–64, 1967
13. Wybenga DR, Ploggi VJ, Dirstine PH *et al*: A direct manual determination of serum total cholesterol with a single stable reagent. **Clin Chem** 16: 980–984, 1970
14. Rowe JW, Andres R, Tobin JD, *et al*: The effect of age on creatinine clearance in men: a cross-sectional and longitudinal study. **J Gerontol** 31: 155–163, 1976
15. McGandy RB, Barrows CH Jr, Spanias A, *et al*: Nutrient intakes and energy expenditure in men in different ages. **J Gerontol** 21: 581–587, 1966
16. Elahi V, Elahi D, Andres R, *et al*: A method of separating aging, cohort, and time effects, and its application to a longitudinal study of nutritional intake in men. Submitted for publication.
17. Hollister LE, Overall JE, Snow HL: Relationship of obesity to serum triglycerides, cholesterol, and uric acid, and to plasma-glucose levels. **Am J Clin Nutr** 20: 777–782, 1967
18. Walker WJ, Lawry EY, Love DE, *et al*: Effect of weight reduction and caloric balance on serum lipoprotein and cholesterol levels. **Am J Med** 14: 654–664, 1953
19. Montoye HJ, Van Huss WB, Brewer WD, *et al*: The effects of exercise on blood cholesterol in middle-aged men. **Am J Clin Nutr** 7: 139–145, 1959
20. Ashley FW Jr, Kannel WB: Relation of weight change to changes in atherogenic traits: the Framingham Study. **J Chron Dis** 27: 103–114, 1974
21. Mann GV: Diet-heart: end of an era. **New Eng J Med** 297: 644–650, 1977
22. Glueck CJ, Mattson F, Bierman EL: Diet and coronary heart disease: another view. **New Eng J Med** 298: 1471–1474, 1978
23. Keys A, Anderson JT, Grande F: Serum cholesterol response to changes in diet (I–IV). **Metabolism** 14: 747–787, 1965
24. Jacobs DR Jr, Anderson JT, Blackburn H: Diet and serum cholesterol: do zero correlations negate the relationship? **Am J Epid** 110: 77–87, 1979
25. Hegsted DM, McGandy RB, Myer ML, *et al*: Quantitative effects of dietary fat on serum cholesterol in man. **Am J Clin Nutr** 17: 281–295, 1965
26. Stulb SC, McDonough JR, Greenberg BG, *et al*: The relationship of nutrient intake and exercise to serum cholesterol levels in white males in Evans County, Georgia. **Am J Clin Nutr** 16: 238–242, 1965
27. Kannel WB, Gordon T: **The Framingham Diet Study: Diet and the Regulation of Serum Cholesterol** (Sect 24). Washington, DC, Department of Health, Education and Welfare, 1970
28. Nichols AB, Ravenscroft C, Lamphiear DE, *et al*: Dietary nutritional intake and serum lipid levels. The Tecumseh Study. **Am J Clin Nutr** 29: 1384–1392, 1976
29. Keys A: Coronary heart disease in seven countries. **Circulation** 41 (Suppl 1): 1–211, 1970
30. Montoye HJ, Block WD, Metzner HL, *et al*: Habitual physical activity and serum lipids: males 16–64 in a total community. **J Chron Dis** 29: 697–709, 1976

Field, T.M.; Huston, A.; Quay, H.C.; Troll, L., and Finley, G.E. (Eds.),
Review of Human Development.
New York: Wiley, 1982

AGING, THE LIFE COURSE, AND MODELS OF PERSONALITY

Robert R. McCrae
Senior Staff Fellow

Paul T. Costa, Jr.
Chief

Section on Stress and Coping
Laboratory of Behavioral Sciences
Gerontology Research Center
National Institute on Aging, NIH

As researchers in the field of personality and aging, we are usually asked only one question: What happens to personality with age? When we answer that the best evidence to date suggests that personality is basically stable and unchanging across the adult portion of the life span, we immediately lose the interest and sometimes the goodwill of the questioner. In our view, such a reaction is prompted by a rather limited conception of the many ways in which the relations between age and personality can be profitably studied. This chapter is an attempt to review briefly a number of alternative models of personality in relation to life-span development and to indicate some of the directions in which one of them, the dimensional continuity model, can be taken.

ALTERNATIVE MODELS OF PERSONALITY AND AGING

The Growth/Decline Model

The most common conception of aging research can be labeled the *growth/decline model.* In this approach, researchers attempt to chart the increase or decrease of a variable as a function of chronological age. Field independence (or spatial ability), for example, is known to increase with age during childhood and adolescence and to decrease thereafter, especially in old age (Schwartz & Karp, 1967). Many physical functions, such as height, weight, and pulmonary capacity, show the same pattern. The major concerns within this model are the identification of variables which show some age-related pattern of growth or decline; the separation of "maturational" changes from generational differences or historical changes; and, ideally, the discovery of the mechanisms which account for the life-span changes in the variable.

A welter of cross-sectional studies of personality variables have implicitly followed this model (Neugarten, 1977), and a small but growing number of longitudinal studies have used repeated measures and cross-sequential techniques to confirm or reject the cross-sectional findings. A recent review (Costa & McCrae, 1980a) argued that the general conclusion to be drawn from these studies was that there are few, if any, meaningful age-related changes in the level of personality traits in the adult years. More precisely, there is no replicated longitudinal evidence of change in such traits as emotional stability, hostility, ascendance, sociability, or imaginativeness. A few personality variables have shown a pattern which can be interpreted as maturational change, but the amount of change is modest, and the range of individual differences at any single age is larger than the variation across ages.

Consider the trait of "masculinity." Studies from projective methods (TAT) have suggested the hypothesis that men decline in some masculine characteristics as they age, while women show an increase in masculine characteristics (Neugarten & Gutmann, 1968). In a study using the Guilford-Zimmerman Temperament Survey on a large sample of men ranging in age from 20 to 90, Douglas and Arenberg (1978) found statistically significant decreases in masculinity which were present in cross-sectional, longitudinal, and cross-sequential analyses. However, the conclusion that aging men become "feminized" (based upon the decreases in the masculinity scores) is premature. In a somewhat whimsical vein, we calculated from the cross-sectional differences that the average man would score the same as the average female college student only if he lived to the age of 269. Extrapolating from longitudinal rates of change, the estimate is 211 years.

To be sure, research to date has not exhausted the personality variables which might be substantially related to age. Within the growth/decline model, the search for personality traits, processes, or structures which are ordered by age continues to remain one direction for research, as does the effort to explain the mechanisms which underlie the small age changes which have already been documented. But the most pervasive domains of personality–neuroticism, extraversion, and openness to experience–show little evidence of age-related change; and an approach which disregards these domains because they do not show such change can hardly claim to provide a comprehensive psychology of aging and personality.

The Life Stage Model

Researchers who employ the growth/decline model have generally been empiricists who prefer to gather facts before propounding theories. But the most elegant models of the interaction of personality and aging have been put forward by stage theorists like Erikson (1950) and, more recently, Levinson et al. (1978). These writers reject the idea that a psychology of adult development can be based on tracing the rise and fall of discrete variables. Instead, they hold that there are qualitative shifts in the nature and relevance of personality variables and syndromes as a function of the stage of adult development. Typically these theorists posit an interaction of social and intrapsychic factors in which personality must be considered both cause and effect in the shaping of the life course.

Erikson describes the life cycle in terms of the succession of stages of psychosocial development. For Erikson, man and society have evolved to a mutual accommodation in which the well-adjusted, mature individual is one whose psychological organization meshes with the social age-grading requirements of his society. Correspondingly, culture has evolved institutions to complement the capacities of the individual at each stage, educating the child concerned with developing industry and revering the elder who has attained integrity.

Levinson et al. (1978) have proposed a complex stage model of adult development in which the life structure, rather than personality itself, is the variable to be explained. Nevertheless, certain personality characteristics are inextricable elements of the life structure and its changes. At the midlife transition, hypothesized to occur universally around age 40, the individual may go through a period of inner turmoil which may resemble neurosis. There is also a reemergence of the repressed wishes and dreams of youth. These intrapsychic changes are instrumental in the reshaping of the life structure for the midlife period.

Such theories are rich and appealing, but their commendable complexity often makes them difficult to test empirically. So interlinked are the elements of individual, society, and history that almost any phenomenon can be explained, though few can be predicted. There are a number of studies which offer support for some of the propositions of Erikson's theory (e.g., Whitbourne & Waterman, 1979), particularly for the period of adolescence and young adulthood, but the theory as a whole cannot be considered empirically established. Furthermore, there have been empirical studies which directly contradict the premises of some stage theories of adult development. An attempt to locate the universal "midlife crisis" (Cooper. 1977) found only a small group of men with signs of a crisis, and these were found to vary in age from 30 to 60, the effective range of the study population. These negative results were confirmed in a second study (Costa & McCrae, 1978). Another recent attempt to detect age-related life stages in respondent's life orientation (futurity) and satisfaction has also found little support for uniform and universal age-related stages (Lacy & Hendricks, 1980).

Critics could question the sensitivity of the measures used in investigating so complex a phenomenon, but such a criticism is itself an admission that the life stage model is at best a difficult basis for the empirical investigation of adult development. More elaborate attempts to address the question of qualitative change in personality at different stages of life using standard personality instruments have occasionally been made. Factor analyses of personality scales within age groups and at different points in time have been used for this purpose but have yielded no evidence of theoretically relevant change in the interrelation of traits. Quite the contrary: Data from large samples of men showed striking invariance of personality structure across age and time (Costa & McCrae, 1980a; McCrae, Costa, & Arenberg, 1980). Again, it must be noted that this is only one of several ways in which qualitative changes might be sought; but there is little in the existing literature to encourage research in this direction.

The basic insight which distinguishes the life stage model theorists from others is their recognition that personality, cultural age norms, and expectations interact in determining the life course of the individual. What is needed is a model which incorporates this conception into a psychology of aging built on sound empirical findings about personality.

The Typological Model

A third model, which makes certain contributions along these lines, might be called the typological model. In this approach, individuals of a given age group are classified into types which are based on the life-style, personality, and adjustment of the individuals. The use of personality measures or ratings and the statistical classification of individuals adds an empirical element to these approaches which is laudable. At the same time, the labels chosen to characterize the groups often summarize holistically the life structure, as well as the personality, of the individual. For example, the "rocking chair men," one of five types identified by Reichard, Livson, and Peterson, (1962) are described as a "passive-dependent group [that] tended to lean on others for material and emotional support. Unambitious men who found little satisfaction in work, they were glad to take it easy when retirement came" (p. 129).

Neugarten, Crotty, and Tobin (1964) used a similar technique to identify six personality types in an aged population. More recently, Maas and Kuypers (1974) again used a typological approach in order to investigate the 40-year predictors of personality and life-style in 70-year-old men and women. They described four life-style clusters for men and six for women, as well as three personality types (based upon the California Q-sort method) for men and four for women. However, Maas and Kuypers found "a relatively random association between personality Q-groups and life style clusters" (p. 156), and thus concluded that personality organization and patterns of life-style are quite independent of each other.

It is the multiplicity of distinct types that is the major shortcoming of the typological approach. If the three studies cited above had concurred in the number and nature of personality or life-style types, there would be a sounder basis for employing this model. As it is, the clustering seems to depend on the particular set of variables included and seems not to be particularly robust across studies. Another point needs to be considered: in all three of these studies, subjects were restricted to older individuals. A cluster analysis of personality and life-style variables which showed that age ordered the clustering of subjects better than gender, occupation, or some other variable would materially strengthen the claim that the various typologies capture the interrelation of age and personality. But within the age range involved, Neugarten et al. (1964) report that "an important finding is that the personality types described here are not, on the whole, related to age. . . . over the wide age ranges from the early fifties to the late eighties" (p. 186).

That different types emerge in different studies should not obscure the fact that all three reach a similar conclusion with regard to the continuity of personality in adulthood. Although they are careful to point out that there is notable change in life-styles, Maas and Kuypers (1974) also show longitudinal evidence of continuity. The 70-year-old men they classify as "unwell-disengaged" were the most explosive, tense, and nervously unstable of the groups 40 years earlier. Likewise, "the fearful-ordering mothers remain, over their adult life course, depressed in mood and activity level, low in adaptive capacity, and low in self worth" (p. 203). Continuity is also a conclusion from the other studies: "The histories of our aging workers suggest that their personality characteristics changed very little throughout their lives"

(Reichard, Livson, & Peterson, 1962, p. 163). "The implication . . . is that personalities maintain their characteristic patterns of organization as individuals move from middle into old age" (Neugarten, Crotty, & Tobin, 1964, p. 187).

The Dimensional Continuity Model

The dimensional continuity model of aging and personality, our preferred model, argues that the cardinal feature of personality in adulthood is the stability of a number of its major dimensions. This premise is based on a growing body of data which show that the retest correlations of personality measures administered over a period of many years are extremely high; in some cases rivaling the short-term retest reliabilities of the measures (Costa & McCrae, 1977; Costa, McCrae, & Arenberg, 1980). Correlations ranging from .59 to .87 over a 12-year interval statistically confirm the impressions of stability which retrospective accounts give and lay the basis for an entirely new approach to the relation of personality to the life course.

It might be useful to compare the present stability model of personality with the epigenetic model of Erikson. Erikson proposes eight stages, each with a better and worse resolution. He argues that success in preceding stages is the best preparation for success in the next but views each new conflict as a potential for change, either positive or negative. Our longitudinal data lead to the simpler, if less optimistic, conclusion that success in any single developmental task is likely to be a reflection of stable personality traits. Rather than the branching life patterns that a series of developmental successes and failures could provide in theory, we would hypothesize a preponderance of straight lines: some individuals would show a constellation of mistrust, doubt, guilt, inferiority, identity diffusion, isolation, stagnation, and despair throughout their lives; others would show a constellation of trust, autonomy, initiative, industry, identity, intimacy, generativity, and integrity; and most people would show a lifelong pattern of moderate adjustment.

The point is that whereas the Eriksonian view leads one to look for the possible emergent changes in personality at each stage, our position emphasizes the continuity in outcomes across the stages. Radical changes in personality may perhaps result from effective therapeutic interventions or catastrophic changes in health or social status, but our data suggest that age, per se, does not bring about stage changes in adult personality.

Unlike typologies, which have not proven easy to replicate, a large literature recently has concurred in the identification of a few basic dimensions of personality traits. Neuroticism and extraversion are found in the theories and measures of Cattell (1973), Guilford (1976), and Eysenck (1960). Another domain, openness to experience (McCrae & Costa, 1980) is beginning to be recognized as a third pervasive dimension of personality (e.g., Tellegen & Atkinson, 1974). Within each of these domains, a number of distinct but covarying traits can be enumerated. Neuroticism includes anxiety, hostility, depression, self-consciousness, impulsiveness, and vulnerability to stress. Extraversion includes warmth, gregariousness, assertiveness, activity, excitement seeking, and positive emotions. Openness is manifested in the areas of fantasy, aesthetics, feelings, actions, ideas, and values. Grouping these traits into three domains provides a model which is reasonably comprehensive (though certainly not exhaustive of personality dimensions) while being conceptually manageable. The demonstrated empirical relations between alternate measures of traits in these domains (Costa & McCrae, 1980a) makes it possible to compare studies using different instruments. The internal consistency, retest reliability, and discriminant validity of these well-constructed objective personality tests also contribute to the conclusion that the dimensional continuity approach offers a sound empirical basis for the study of aging and personality.

What the dimensional continuity approach lacks, in the eyes of most researchers, is any application. Having said that personality is stable, what more can be said? Is there any future to research on personality and adult development, other than to refute or confirm, to qualify or delimit the central claim of stability?

In fact, there are a number of important directions for future research. Elsewhere (Costa & McCrae, 1980a) we have argued that one such direction is the search for an explanation of personality stability. Little is known about the mechanisms which maintain characteristic levels of various dispositions, although a number of theoretically relevant possibilities can be cited, from genetics to self-image. The institutionalization of the self in the interlocking obligations and expectations of the life structure is doubtless another source of personality stability.

This chapter, however, is directed at another way in which personality and aging can fruitfully be studied. Here we will argue that the enduring dimensions of personality can be viewed as a framework for understanding the life course of the aging person. Rather than taking personality as the dependent variable to be understood in terms of age, or stage in some career, we view personality dimensions as independent variables which function jointly with age and stage to influence some of the outcomes of life. In looking for these interrelations of personality and the life course, we will be following in the tradition of major personality theorists like Murray (1938) who held that the phenomenon to be explained is not a single process or specific behavior, but the complete life of the individual (White, 1963).

THE LIFE COURSE AS A FUNCTION OF AGE AND PERSONALITY

Life-span developmentalists share a concern for the course of life as a major element in their theories and research. They differ, however, in their goals. Some researchers are primarily interested in psychological or biological processes and view aging as a quasi manipulation, the effects of which may help elucidate the mechanisms underlying the process. Sociologists (e.g., Elder & Rockwell, 1974) take social, historical, and developmental "ages" into account in attempting to understand career development or such social phenomena as the unwed mother. Personality researchers have historically considered their primary concern to be an understanding of the whole person, both at a given time and across the life span; and some, like Erikson (1962), have written biographies. The relevance of enduring personality dispositions to the work of life-span developmentalists will necessarily vary with their goals. Our purpose here is to describe one of the several approaches to the study of the life span, the approach we find valuable for the study of aging and personality.

Determinants of the Life Course

It is possible to view a person's life in cross section and to describe what Levinson et al. have called a "life structure." At any given time the life structure can be described in terms of occupational, social, and family roles; intimate, personal, and professional relationships; and the goals, values, motives, and memories which constitute the inner aspect of the life structure. One of the tasks of the individual is to manage all these elements at one time, to avoid role conflict, and to accommodate all the intrapsychic needs, values, and preferred styles. Murray and Kluckhohn (1953) refer to this process as "scheduling."

Viewed longitudinally, the sequence of any one of these elements is often called a *career*, and it is possible to speak of social, leisure time, or family as well as occupational careers. From the psychological side, Murray calls these temporal sequences "serials." Taken as a whole, the more-or-less coherent progress of all these aspects of life can be called a "life course," which is the proper subject of a complete biography.

The life-span developmentalist is likely to view the life course as a series of *changes* which are shaped by the succession of age-related roles prescribed by the culture, by the biological and cognitive development of the individual, and by the particular historical events, shared and idiosyncratic, that define the context in which the individual ages. The choice of retirement, for example, can be seen to be influenced by social policies of mandatory retirement, by the physical health of the individual, and by prevailing economic conditions. As an element in the life course, retirement then influences a host of other events and provides the individual with new choices for the use of leisure time, new residence, and so on.

By and large, developmentalists have concentrated on the explanation for changes in the life course and have given less thought to factors which provide continuity. For the individual, however, continuity is as important as change, for it provides the basis for a sense of identity. A number of sources of continuity in the life course can be easily recognized. The social structures of the family and social class provide an enduring set of opportunities and expectations while biological and cognitive abilities of the individual set certain stable limits to achievement. This chapter will argue that the stability of dimensions of personality is yet another powerful source of continuity in the life course of the individual, although it has been perhaps less widely recognized as such by life-span theorists.

Both continuity and ordered change are necessary for the smooth functioning of society, and mechanisms for preserving continuity (such as marriage contracts and seniority systems) have been institutionalized. But the student of personality must also point out that much of life's continuity is the result of the individual's own action. Most people have considerable say in the shaping of their own lives, and the successful management of the life structure at any one time depends on a history of preparation and planning. In making these decisions, one of the major considerations of the individual is his or her own personality dispositions. Indeed, if personality were not stable, our ability to make wise choices about our future lives would be severely limited.

Personality and Critical Life Choices

The statement that personality and the life course are "interrelated" means that there are a number of phenomena that are best explained in terms of both conceptions jointly. In Runyan's (1980) terms, we could say that personality influences the state of the individual at any given stage. Some of the examples which we will provide are obvious on reflection, and some are empirical discoveries. It is also relatively easy to generate speculations about possible relations to be confirmed or disconfirmed by research. For this reason, we believe this to be a fruitful model for developmental research.

Perhaps the most obvious example of the interaction of aging and personality is to be found in the transitions of the life course. On the basis of biological capacity and social requirements, cultures have dictated that certain roles must be adopted or given up at certain times. Young adults are expected to marry, begin a family, and take up an occupation. Middle-aged persons are supposed to sustain their families and advance their careers. Older individuals, at least until recently, were expected to step down from positions of responsibility and adapt gracefully to a period of relative inactivity. But at each transition there is also choice, and in a culture like ours, where the individual is given wide latitude in the choice of roles, the role of choice becomes more central in the shaping of the life course.

Most individuals start a career in their 20s. But what determines the selection of a particular occupation? Intelligence, education, social class, role models, and a large element of chance go into the choice, but so do personality dispositions. Holland (1966) has developed a theory of occupational development based on personality types, and scales of occupational interest are known to correlate with dimensions of personality, particularly extraversion (Costa, Fozard, & McCrae, 1977). Young people may not know the professions most suited to their temperaments, but if they have made a mistake, they soon discover it. The period of occupational adjustment in the 20s is in part a period of self-discovery in which the tastes, interests, and capacities of the individual become more apparent to him or her. The introvert does not last long as a door-to-door salesperson and is not likely to try that line of work again.

Recently, attention has been drawn to the phenomenon of midlife career shifts (Clopton, 1973). To the extent that these changes are voluntary, they highlight the proactive choice of the individual in shaping the life course. From the viewpoint of personality stability, we might form two hypotheses about the kinds of people who would choose a new career. First, we would expect that individuals whose initial career choice was incompatible with their temperament would be most likely to change, in an attempt to find a more satisfying occupation. There is some evidence from the study of vocational interests that this process does occur (Strong, 1955). Second, we might hypothesize that the need for change or variety is greater in some individuals and that after a number of years in a particular field, they might want to move on. Our own research provides some data consistent with this hypothesis (Costa & McCrae, 1978). Men who had changed their line of work in the previous 10 years were significantly higher in openness to experience than those who had not. It is possible, of course, that the change in vocation led to higher openness instead of the other way around, and longitudinal research is currently in progress which would allow an assessment of that alternative.

Both vocational counseling and industrial selection have capitalized on the association between personality and occupation in the use of interest and attitude scales. Less attention has been paid to the role of personality in determining the use of time after retirement, although there is every reason to believe that personality should figure prominently. Freed from the necessity of working, the older person can spend time as he or she wishes, and an even clearer expression of individual temperament should result.

The continuity of abilities, interests, and values into old age can result in a continuity of activities. Havighurst et al. (1979) report a study on the postretirement publications of a group of scientists. They find clear evidence that retirement per se has little effect and that there is a pattern of continuity between productivity before and after the event. Despite major changes in social-role requirements, the majority of their subjects continued to publish at about the same rate.

Scientists, however, belong to a rather small group who can, if they desire, continue to work after formal retirement. For most people, retirement is a source of discontinuity, but it is also an opportunity to do other things which may have been impossible before. Sociologists have dealt extensively with the variables which affect the age of retirement, including health, income, and job satisfaction. But relatively little research has gone into a specification of the determinants of how retired people employ their own leisure. What kinds of people travel? What kinds retire to a farm? Who joins senior citizen clubs? Who becomes a burden on the local health clinic? Who goes back to college? Any complete theory of aging and the life course must surely address these questions, and it is probable that stable personality traits will be one of the explanatory factors.

Personality and Adaptation

In addition to the choice of roles and relationships, personality influences adjustment to the circumstances—chosen or not—in which the individual finds himself. Any clinical psychologist or psychiatrist who has taken life histories knows that the maladjusted adult typically shows a lifelong pattern of poor adaptation. Prospective studies from deviant children confirm this impression (Robins, 1966), as do recidivism rates for treated patients of all kinds (Moss & Susman, 1980).

Research on the "midlife crisis," also, is instructive here (Costa & McCrae, 1978). A series of questions was asked which addressed the characteristics and concerns of men in the middle portion of their lives: questions about marital satisfaction, declining power and potency, career fulfillment, and problems with children and with aging parents. Although there was no evidence that individuals scoring high on this checklist of problems clustered at any particular age within the range from 30 to 60, there was a strong association ($r = .51$) with concurrent measures of neuroticism. Further, personality scales measuring neuroticism 10 years earlier showed a highly significant predictive correlation. Those individuals with a history of neurotic traits were most likely to suffer during midlife the complaints identified as constituting a midlife crisis. In a similar vein, Lowenthal and Chiriboga (1972) have reported that most women do not experience unhappiness in the "empty nest" period and that those who do have a preexisting history of maladjustment.

The 40-year longitudinal findings of Maas and Kuypers (1974) give a similar impression with regard to old age. They argue that "old age does not usher in or introduce decremental psychological processes. Rather, old age may demonstrate, in perhaps exacerbated forms, problems that have long-term antecedents" (p. 203). Research on health complaints (Costa & McCrae, 1980b) similarly finds no longitudinal increase in symptoms for most body systems, but more neurotic individuals of all ages report more physical problems. The same pattern of invariant relations between personality and outcome variables across the adult life span is seen in studies on psychological well-being or personal adjustment to aging (Costa, Mc-

Crae, & Norris, 1980). Even death anxiety, which might be imagined to be most relevant to older persons, shows little relation to age, but at all ages, it is the characteristically anxious person who shows fear of death (Kastenbaum & Costa, 1977).

Questions and Methods for Future Research

One of the chief obstacles to the study of lives is the sheer quantity of information which must be synthesized in order to make sense of the subject. Biographers adopt some implicit framework, and their task is simplified by the fact that they treat one individual and are allowed post hoc and idiographic interpretations. The life-span developmentalist who desires a scientific theory of the course of human life likewise must account for the similarities and differences of all people. Clearly some organizing principle is necessary, and the continuity of personality dimensions might well provide this principle and form a better basis for an understanding of the life course. Some empirical observations can be profitably understood in this framework, but the extent and limits of the approach are unknown. For the researcher interested in exploring the life course from the viewpoint of the individual, we suggest such questions as these: How do the lives of introverts differ from the lives of extraverts? Which aspects of the life structure are influenced by openness to experience? Does neuroticism or the poor coping styles associated with it interfere in the individual's ability to make effective schedules and serials to order and plan a life? What role do personality dispositions play in adapting to stressful life events?

Another difficulty in the study of lives is the time required to observe the phenomenon. Historical biography is one solution to this problem, as is the analysis of personal documents covering an extended time (Allport, 1965). Prospective longitudinal studies are invaluable for the objectivity and pertinence of the data they provide. But researchers should also utilize the retrospective account of the individual as a source of data. Older persons in particular have a unique perspective on the life course, and through the process of reminiscence, many of them are engaged in making sense of their own lives and of transmitting their insights to others (Butler, 1963). Thus far, retrospective accounts and prospective longitudinal studies have shown substantial agreements in pointing to stability in personality and continuity in life course. The accumulated experience of older men and women should be regarded as a scientific, as well as a social, resource.

CONCLUSIONS

Personality has usually been regarded as an outcome of living, thought to develop or decline with age, or to metamorphose through different life stages. But objective personality data and clinical impressions find more stability than change in the period of adulthood and old age.

This fact encourages a recasting of the fundamental question–instead of asking how personality is changed by aging, we now can ask how the life course is shaped by enduring personality dispositions. The widely recognized and replicated dimensions of neuroticism, extraversion, and openness to experience provide a useful framework in which to organize data on continuity and changes in the life structure. Stable personality dimensions can help explain the choices which are required at certain age-related transition points, the maintenance of life-style across different developmental periods, and the level of adaptation at all ages.

Universal developmental stages cannot account for the continuity of the life course, just as enduring individual differences cannot account for all important life changes. But a model that considers biological, cultural, and historical influences on aging in conjunction with stable dimensions of personality offers promise as a sound basis for an understanding of the life course.

REFERENCES

Allport, G. W. (Ed.) *Letters from Jenny*. New York: Harcourt, Brace & World, 1965.

Butler, R. N. The life review: An interpretation of reminiscence in the aged. *Psychiatry*, 1963, **26**, 65–76.

Cattell, R. B. *Personality and mood by questionnaire*. San Francisco: Jossey-Bass, 1973.

Clopton, W. Personality and career change. *Industrial Gerontology*, 1973, 9–17.

Cooper, M. W. An empirical investigation of the male midlife period: A descriptive, cohort study. Unpublished manuscript, University of Massachusetts at Boston, 1977.

Costa, P. T., Jr., Fozard, J. L., & McCrae, R. R. Personological interpretation of factors from the Strong Vocational Interest Blank scales. *Journal of Vocational Behavior*, 1977, **10**, 231–243.

Costa, P. T., Jr., & McCrae, R. R. Age differences in personality structure revisited: Studies in validity, stability, and change. *Aging and Human Development*, 1977, **8**, 261–275.

Costa, P. T., Jr., & McCrae, R. R. Objective personality assessment. In M. Storandt, I. C. Siegler, & M. F. Elias (Eds.), *The clinical psychology of aging*. New York: Plenum, 1978.

Costa, P. T., Jr., & McCrae, R. R. Still stable after all these years: Personality as a key to some issues in adulthood and old age. In P. B. Baltes & O. G. Brim (Eds.), *Life span development and behavior* (Vol. III). New York: Academic, 1980. (a)

Costa, P. T., Jr., & McCrae, R. R. Somatic complaints in males as a function of age and neuroticism: A longitudinal analysis. *Journal of Behavioral Medicine*, 1980, **3**, 245–257. (b)

Costa, P. T., Jr., McCrae, R. R., & Arenberg, D. Enduring dispositions in adult males. *Journal of Personality and Social Psychology*, 1980, **38**, 793–800.

Costa, P. T., Jr., McCrae, R. R., & Norris, A. H. Personal adjustment to aging: Longitudinal prediction from neuroticism and extraversion. *Journal of Gerontology*, 1980, **36**, 78–85.

Douglas, K., & Arenberg, D. Age changes, cohort differences, and cultural change on the Guilford-Zimmerman Temperament Survey. *Journal of Gerontology*, 1978, **33**, 737–747.

Elder, G. H., & Rockwell, R. C. The life course and human development: An ecological perspective. *International Journal of Behavioral Development*, 1979, **2**, 1–33.

Erikson, E. H. *Childhood and society*. New York: Norton, 1950.

Erikson, E. H. *Young man Luther: A study in psychoanalysis and history*. New York: Norton, 1962.

Eysenck, H. J. *The structure of human personality*. London: Methuen, 1960.

Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. *The Guilford-Zimmerman temperament survey handbook: Twenty-five years of research and application*. San Diego, Calif.: Edits, 1976.

Havighurst, R. J., McDonald, W. J., Maeulen, L., & Mazel, J. Male social scientists: Lives after sixty. *The Gerontologist*, 1979, **19**, 55–60.

Holland, J. L. *The psychology of vocational choice: A theory of personality types and model environments*. Waltham, Mass.: Blaisdell, 1966.

Kastenbaum, R., & Costa, P. T., Jr. Psychological perspectives on death. In M. R. Rosenweig & L. W. Porter (Eds.), *Annual Review of Psychology*, 1977, **28**, 255–249.

Lacy, W. B., & Hendricks, J. Developmental models of adult life: Myth or reality. *International Journal of Aging and Human Development*, 1980, **11**, 89–110.

Levinson, D. J., Darrow, C. N., Klein, E. B., Levinson, M. H., & McKee, B. *The seasons of a man's life*. New York: Knopf, 1978.

Lowenthal, M. F., & Chiriboga, D. Transition to the empty nest. *Archives of General Psychiatry*, 1972, **26**, 8–14.

Maas, H. S., & Kuypers, J. A. *From thirty to seventy*. San Francisco: Jossey-Bass, 1974.

McCrae, R. R., & Costa, P. T., Jr. Openness to experience and ego level in Loevinger's Sentence Completion Test: Dispositional contributions to developmental models of personality. *Journal of Personality and Social Psychology*, 1980, **38**, 1179–1190.

McCrae, R. R., Costa, P. T., Jr., & Arenberg, D. Constancy of adult personality structure in males: Longitudinal, cross-sectional and times of measurement analyses. *Journal of Gerontology*, 1980, **35**, 877–883.

Moss, H. A., & Susman, E. J. Constancy and change in personality development. In O. G. Brim, Jr., & J. Kagan (Eds.), *Constancy and change in human development*. Cambridge, Mass.: University Press, 1980.

Murray, H. A. *Explorations in personality*. New York: Oxford, 1938.

Murray, H. A., & Kluckhohn, C. Outline of a conception of personality. In C. Kluckhohn & H. A. Murray (Eds.), *Personality in nature, society, and culture* (2nd ed.). New York: Knopf, 1953.

Neugarten, B. L. Personality and aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold, 1977.

Neugarten, B. L., Crotty, W. J., & Tobin, S. Personality types in an aged population. In B. L. Neugarten (Ed.), *Personality in middle and later life*. New York: Atherton, 1964.

Neugarten, B. L., & Gutmann, D. L. Age-sex roles and personality in middle age: A thematic apperception study. In B. L. Neugarten (Ed.), *Middle age and aging*. Chicago: University of Chicago Press, 1968.

Reichard, S., Livson, F., & Peterson, P.G. *Aging and personality*. New York: Wiley, 1962.

Robins, L. N. *Deviant children grow up*. Baltimore: Williams and Wilkins, 1966.

Runyan, W. McK. A stage-state analysis of the life course. *Journal of Personality and Social Psychology*, 1980, **38**, 951–962.

Schwartz, D. W., & Karp, S. A. Field dependence in a geriatric population. *Perceptual and Motor Skills*, 1967, **24**, 495–504.

Strong, E. K., Jr. *Vocational interests 18 years after college*. Minneapolis, University of Minnesota, 1955.

Tellegen, A., & Atkinson, G. Openness to absorbing and self-altering experience ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, 1974, **83**, 268–277.

Whitbourne, S. K., & Waterman, A. S. Psychological development during the adult years: Age and cohort comparisons. *Developmental Psychology*, 1979, **15**, 373–378.

White, R. W. (Ed.) *The study of lives: Essays on personality in honor of Henry A. Murray*. New York: Atherton, 1963.

Journal of Gerontology
1980, Vol 35, No 6, 877-883

Constancy of Adult Personality Structure in Males: Longitudinal, Cross-sectional and Times-of-Measurement Analyses[1]

Robert R. McCrae, PhD, Paul T. Costa, Jr., PhD,
and David Arenberg, PhD[2]

Constancy or change in adult personality organization can be assessed by comparing the factor structure of personality instruments at different ages, and some studies have reported cross-sectional differences in structure. The present study compares the factor structure of the Guilford-Zimmerman Temperament Survey scales longitudinally in three administrations 6 years apart and cross-sectionally in three age cohorts. Additional analyses compare first administration data collected in two successive decades to test for variation in structure resulting from cultural change over that period. Subjects were 769 male volunteers in the Baltimore Longitudinal Study of Aging, aged 17 to 97 at the time of the first administration. Three varimax-rotated principal components were extracted in each of eight analyses. Results show no systematic evidence of variation in structure in any of these groups, with coefficients of factor congruence ranging from .83 to .99. This longitudinally demonstrated invariance of personality structure is discussed in terms of the implications for the stability of personality organization throughout the adult years and for the use of personality tests in elderly groups.

Key Words: Personality structure, Longitudinal, Factor analysis, Adult males, Stability.

IN recent years the issue of constancy or change in personality through the adult portion of the lifespan has become a major focus of attention (Neugarten, 1977). Indeed, a central question which faces adult developmentalists is whether a stability, ordered change or what Gergen (1977) calls an "aleatory change" (chance) model best describes personality in the adult years. Gergen notes a "general antagonism toward the stability orientation," and some theorists (McCall, 1977) argue that the study of development is necessarily the study of change. Within a life-span perspective, however, change in childhood and stability in adulthood may both be part of the developmental course of a variable.

The stability position has recently received support from longitudinal studies (Costa & McCrae, 1977; Costa et al., 1980) that show that the relative standing of individuals on personality dimensions is fairly constant over time. However, test-retest correlational methods, by which the stability model can be tested, are incapable of demonstrating ordered change. In contrast, longitudinal analyses of the mean levels of personality variables may give evidence of either constancy or ordered change (e.g., maturational decline). Recent longitudinal analyses using standard measures, large samples, and extensive time intervals concur in finding either small changes or no changes at all in mean levels of personality traits (Costa & McCrae, 1978; Douglas & Arenberg, 1978; Siegler et al., 1979). Neither monotonic change with age nor the curvilinear trends predictable from stage conceptions of adult life were typically found in these studies.

A third way in which meaningful change in personality can occur across the life-span involves reorganization in the relations among traits, a change that could be assessed through factor analysis. Qualitative changes in the organization of personality might be manifest as changes in the number or composition of factors or dimensions of personality traits.

---

[1] Portions of this paper were presented at the Gerontological Society meetings, Nov., 1979, Washington, DC. Requests for reprints should be sent to Robert R. McCrae, PhD, Section on Stress and Coping, NIA Gerontology Research Ctr., Baltimore City Hospitals, Baltimore, MD 21224

[2] Gerontology Research Ctr. (Baltimore), NIA Dept. of Health, Education and Welfare, Bethesda and the Baltimore City Hospitals, Baltimore, MD 21224

Personality *differentiation*, a hallmark of development, may be seen in a larger number of distinguishable factors. Different organizations of personality variables, different "syndromes," may appear with age, as traits take on new meanings or new functions and significance. A number of theoretical positions lead to the expectation of such qualitative changes. Jung's process of individuation postulates a balancing of opposed tendencies that may imply the attenuation of bipolar traits. Maddi's (1976) activation theory proposes a continuing process of differentiation and integration throughout the life-span, and Neugarten's (1964) concept of interiority involves "increased consistency and decreased complexity."

Changing factor structure would have methodological, as well as theoretical, implications for gerontologists. If major differences were found in the factor structure of standard personality tests in older samples, the use and interpretation of the tests in these groups could be challenged. Some tests, including the Sixteen Personality Factor Questionnaire (16PF; Cattell et al., 1970) and the Guilford-Zimmerman Temperament Survey (Guilford et al., 1976), relied explicitly on factor analytic techniques in their development and require age-invariance to maintain their "factorial validity." All tests must demonstrate their validity by consistent patterns of convergent and divergent relations. If these relations show a marked change with age the meaning of the scales and perhaps the construct itself has altered somehow. Essentially, all previous work on test development (except that conducted on elderly samples) would be called into question for elderly populations, as would all research that has employed these tests in the study of aging.

Age-related differences in the factor structure of personality tests have been reported. A cross-sectional examination (Costa & McCrae, 1976) of the cluster structure of the 16PF showed age invariance in the first two clusters, identified as Anxiety-adjustment (or Neuroticism) and Extraversion. The composition of the third cluster, which was interpreted as Openness to Experience, appeared to show age-related differences. However, this age-related variation was not found in a longitudinal replication (Costa & McCrae, 1977), nor was there any evidence of age difference in the structure of Openness to Experience when alternate measures of the component traits were employed (Costa & McCrae, 1980).

Like the 16PF, the Guilford-Zimmerman Temperament Survey (GZTS) is a factor-based personality questionnaire that has been used extensively in research and counseling. It is intended to measure a variety of normal personality traits including cognitive, emotional, and interpersonal styles. A three-factor structure of the 10 scales has been confirmed in over 20 analyses (Guilford et al., 1976). In one of these, using a sample of men age 20 to 60, Bendig (1960) reported the usual three factors, but he found that two scales, Emotional Stability and Masculinity, showed different factor loadings when factor structures in 10-year cohorts were compared. Among 20 and 30 year olds, both these scales formed part of the "Friendliness"[3] or "Emotional Health" factor. However, among 40 and 50 year olds, Emotional Stability loaded on a "Social Activity" factor, whereas Masculinity showed no high loadings on any of the three factors. Bendig speculated that these differences in structure were related to differences in the mean levels of the variables, particularly Masculinity. He acknowledged, however, that his cross-sectional study could not demonstrate that these differences resulted from maturation.

Recent evidence on the relative constancy of mean level in the GZTS (Douglas & Arenberg, 1978) casts doubt on this interpretation and makes a replication of the Bendig findings timely. The present paper reexamines the question of age-related differences in the structure of personality cross-sectionally by comparing the factor structure of GZTS scales in different cohorts and longitudinally by comparing the first with the second and third administrations, 6 and 12 years later.

Schaie (1965) has argued that with the passage of historical time, cultural changes may alter the stimulus value of test items; tests may become "dated." This reasoning, applied to critiques of conventional longitudinal repeated measures analyses, applies with equal force to structural considerations. In an attempt to determine whether cultural changes affect the organization of personality variables, the present study also compares the factor structure of the GZTS for men who first completed the survey in the 10 year periods before and after July, 1968.

---

[3] Note that the label "Friendliness" applied either to the scale or factor is potentially misleading. Individuals low on this scale show hostile and uncooperative interpersonal styles. High scorers are not hostile, but neither are they necessarily warm or sociable. In fact, the correlation between the Friendliness and Sociability scales is .02 (N = 769) in our sample.

METHOD

Subjects. — Participants in the Baltimore Longitudinal Study are a highly select, generally healthy group of male volunteers, 96% white, 88% married, who have agreed to return for testing at fixed intervals. The majority (80%) work in, or are retired from, scientific, professional, or managerial positions. Most (93%) are high school graduates, and 71% are college graduates. Data in this paper are from responses of participants who entered the study from late 1958 through 1978. At the time of first administration of the GZTS, age ranged from 17 to 97 (N = 769, mean = 49.8 years).

Second administration data were obtained from 346 men aged 25 to 91 (mean = 57.6 years); third administration data were from 171 men aged 33 to 86 (mean = 61.9 years). Smaller sample sizes result from varying number of years in the study as well as death and withdrawal from the study. Subjects who returned for second or third administration tended to be higher than non-repeats in Emotional Stability, Objectivity, Friendliness and Personal Relations, and lower in Ascendance (Douglas & Arenberg, 1978).

Procedure. — Each subject was given the standard GZTS instructions individually and completed the questionnaire during the remainder of his 3-day visit to the Gerontology Research Center. For each item, subjects chose "Yes" or "No" or "?". Each scale consists of 30 items, but only "Yes" and "No" responses contribute to the total score. A measure was invalidated for any scale with more than three "?" responses, a procedure suggested by Guilford and Zimmerman (1949). Subjects were readministered the GZTS approximately every 6 years. Because of complications in scheduling, a few subjects took the test two years in succession, or failed to take the first retest but did take the second. To maintain consistency of the time interval and number of administrations, longitudinal analyses are limited to subjects who took their second GZTS 5.0 to 7.9 years after their first (mean = 6.6 years), and to those who took their third GZTS 11.0 to 15.4 years after their first (mean = 12.9 years).

Analyses. — Factor analyses were performed on correlation matrices computed for eight groups. Mixed longitudinal comparisons are based on analyses of first administration data from 769 men, second administration data from a subset of the first group consisting of 346 men, and third administration data from a further subset of 171 men. Cross-sectional comparisons are based on first administration data from three age groups: young (17 to 44, mean age = 34.4, N = 314); middle (45 to 59, mean age = 51.6, N = 242); and old (60-97, mean age = 70.4, N = 213). To assess possible structural differences resulting from time-of-measurement effects, the sample was divided into two groups: 455 men who first took the GZTS before July, 1968, (age range = 17 to 83, mean = 52.1), and 314 men who first took the GZTS after that date (age range = 18 to 96, mean = 45.6). To avoid possible confounding with practice effects, these analyses were limited to first administration data. All factor analyses were restricted to subjects with valid scores on all 10 GZTS scales.

RESULTS

Both principal components and principal axes factor analyses were examined. Three factors had eigenvalues greater than unity and accounted for similar amounts of variance in all eight analyses: 28.3% to 30.3% for the first factor; 20.8% to 22.6% for the second; and 11.8% to 13.7% for the third. After varimax rotation comparison of the two methods of factoring showed highly similar results. Only the principal components solutions are presented, because these results are somewhat clearer.

Table 1 shows factor loadings across analyses of data from three administrations, three age groups, and two times-of-measurement. The first factor, which might best be seen as a general Neuroticism factor, has been labeled "Emotional Health" (Guilford et al., 1976) and the structure shown corresponds well with other reported analyses. Emotional Stability, Objectivity, Friendliness, Personal Relations, and Masculinity are consistent definers of this factor; low Thoughtfulness is marginal. The post-1968 analysis shows small contri-

Table 1. Factor Loadings for GZTS Scales Across Administrations, Age Groups, and Times-of-Measurement

	Longitudinal Administrations			Cross-Sectional Age Groups			Times-of-Measurement	
	1st (N = 769)	2nd (N = 346)	3rd (N = 171)	17-45 (N = 314)	46-59 (N = 242)	60-97 (N = 213)	Pre-1968 (N = 455)	Post-1968 (N = 314)
	Factor I. Emotional Health vs Neuroticism							
General Activity	-12	-14	-03	-15	-06	-16	10	-16
Restraint	28	18	16	22	27	23	23	33
Ascendance	07	09	13	07	09	01	06	13
Sociability	22	17	12	22	16	21	12	32
Emotional Stability	73	73	72	67	76	73	76	69
Objectivity	86	84	85	85	87	87	87	84
Friendliness	77	76	72	78	76	76	74	79
Thoughtfulness	-25	-31	-36	-33	-24	-34	-30	-19
Personal Relations	73	71	67	72	71	72	74	71
Masculinity	46	50	57	56	48	49	46	44
	Factor II. Social Activity or Extraversion							
General Activity	68	68	76	70	66	63	66	72
Restraint	-33	-36	-34	-19	-32	37	-31	-34
Ascendence	86	88	86	85	87	84	86	87
Sociability	80	78	77	81	82	77	80	79
Emotional Stability	30	30	38	40	27	31	31	28
Objectivity	17	25	24	23	19	14	17	17
Friendliness	-30	-31	-36	-24	-28	-23	-31	-28
Thoughtfulness	21	19	23	16	11	14	16	07
Personal Relations	04	-01	03	13	06	-01	-01	14
Masculinity	-08	-05	00	-11	-10	-12	-09	-06
	Factor III. "Thinking Introversion"							
General Activity	-10	-09	02	-11	-03	-15	-03	-17
Restraint	71	76	79	79	69	73	76	67
Ascendence	07	05	07	18	-04	07	05	10
Sociability	-04	06	-31	-09	-02	-05	-10	-02
Emotional Stability	-13	-10	-09	-17	-12	04	-01	-24
Objectivity	-12	-07	-10	-09	-07	-03	-04	-22
Friendliness	15	12	-07	13	23	11	14	12
Thoughtfulness	81	80	74	79	78	82	78	84
Personal Relations	08	03	-17	15	03	-07	00	11
Masculinity	-33	-10	17	-13	-32	-12	-21	-44

Note: Varimax-rotated principal components. Decimal points omitted.

butions from Restraint and Sociability. No differences appear between cohorts on the loadings of the two scales Bendig found to vary: Emotional Stability and Masculinity. Both maintain similar loadings in all age groups. The inclusion of low Masculinity among the definers of Neuroticism is somewhat unusual; it may result either from the use of an exclusively male sample or from the fact that 10 of the 30 items in this scale concern susceptibility to the emotions of fear and disgust.

The pattern of General Activity, Ascendance, and Sociability, seen in the second factor in all eight analyses, has been labeled "Social Activity" by Guilford, but could also be identified as social Extraversion. Emotional Stability, low Restraint, and occasionally low Friendliness show small contributions to this factor.

The third factor has been designated "Thinking Introversion" by Guilford, but is not to be confused with the Introversion-Extraversion factors of Eysenck (1960) or Cattell (Cattell et al., 1970). Clearly composed of Restraint and Thoughtfulness across all eight analyses, it shows a small contribution from low Masculinity in some cases. On psychological grounds the meaningfulness of the third factor is questionable. Typically, other personality tests do not yield such a factor, nor do Guilford et al. (1976) discuss any clinical significance or counseling application of the factor. As a measure of sensation-seeking or impulsivity, Restraint may more properly belong in the domain of Extraversion, and small but consistent negative loadings of Restraint on the second factor are consistent with this hypothesis.

Maturational changes in personality structure should appear in both longitudinal and cross-sectional comparisons. As is clear from Table 1, the major definers are the same at each time and in each age group. Small variations in loadings do occur, but they do not show a clear direction or pattern replicated across longitudinal and cross-sectional analyses.

To quantify these impressions of invariance, coefficients of factor congruence (Gorsuch, 1974) were calculated between corresponding factors for administrations, age groups, and times-of-measurement. Used with principal components, as they are here, these coefficients are equivalent to the product moment correlations between factor scores. Table 2 gives the coefficients of congruence for corresponding factors. All coefficients are above .98 for the first two factors; for the third and smallest factor, they range from .83 to .98.

Table 2. Coefficients of Congruence for Corresponding Factors in Different Administrations, Age Groups, and Times-of-Measurement.

	Factors		
	I	II	III
Administrations			
First vs Second	.98	.99	.98
First vs Third	.99	.99	.83
Second vs Third	.99	.99	.91
Age Groups			
Young vs Middle	.99	.99	.95
Young vs Old	.99	.98	.95
Middle vs Old	.99	.99	.95
Times-of-Measurement			
Pre vs Post-1968	.99	.99	.93

Additional analyses were conducted on data from the second administration for subjects aged 25 to 45 ($N = 60$), 46 to 59 ($N = 154$), and 60 to 91 ($N = 132$); and on data from the third administration for subjects aged 33 to 62, ($N = 84$) and 63 to 86 ($N = 87$). In all five analyses, three factors had eigenvalues above 1.0, and, despite small sample sizes, generally similar structures were observed. These are reflected in congruence coefficients (when compared with the full first administration solution) ranging from .96 to .99 for the Neuroticism factor, .91 to .99 for the Extraversion factor, and .58 to .99 for the "Thinking Introversion" factor.

Finally, in order to parallel traditional longitudinal designs in which the same subjects are tested on successive occasions, comparisons were made between different administrations restricting subjects to those with complete data at all administrations. Although the number of subjects meeting this qualification at all three administrations was relatively small ($N = 123$), the same general pattern of definers was replicated at each administration. Only Masculinity varied somewhat, having a larger loading on the "Thinking Introversion" factor (−.41) than on the Neuroticism factor (.25) in the first administration. On the second and third administrations Masculinity was clearly a definer of the first factor. When analyses were conducted for the 324 subjects who had complete data on both first and second administrations, coefficients of congruence for the corresponding factors across the two administrations were .99, .99, and .97.

DISCUSSION

The invariance of factor structures in the GZTS in the present sample is clearly evident. Definers of factors stand out in each case from the marginal elements that show slight variations from one analysis to another. Despite aging, attrition, and possible practice effects, the same pattern is seen at each administration. The only divergence from high factor congruence is found in the case of the third factor at the third administration, when a coefficient of .83 is observed. Although this slight structural variation could result somehow from repeated exposure to the test, it is interpretated more simply as the result of error in the smallest factor and smallest sample. The specific suggestion of Bendig that

Emotional Stability and Masculinity should vanish from the Emotional Health factor in older cohorts is not supported either cross-sectionally or longitudinally, and no other difference or change emerges as a likely hypothesis for further verification.

Sociologically oriented developmentalists (Riley, 1971) note the importance of social and historical change in the life of the individual, and time-of-measurement effects are found in measures of attitudes and opinions and occasionally in measures of personality (Baltes & Nesselroade, 1972). But the basic structure of personality traits appears to be less subject to these influences. In the present data, no meaningful difference can be seen when comparing measurements before and after 1968, yet the decades of the 1960s and 1970s are surely different enough to make a difference if the structure of the GZTS and similar personality tests were particularly sensitive to historical and cultural shifts.

The substantive results of these factor analyses are also noteworthy. As in analyses of the 16PF (Cattell, 1973), the California Psychological Inventory (Mitchell & Pierce-Jones, 1960), and the Eysenck Personality Inventory (Eysenck, 1960), the two largest factors can be identified as Neuroticism and Extraversion. The invariance of structure in Neuroticism and Extraversion is of particular importance, because many of the personality variables of interest to gerontologists — adjustment, depression, anger, activity, vigor, zest — fall within these domains.

These results agree with other recent findings of invariance in personality structure across age groups (Costa & McCrae, 1980) and favor the stability model for objectively measured personality traits throughout adulthood, at least for males. Of course, a number of other ways of defining personality and its organization do not rely on the concepts and measures of the trait approach adopted in this study. Some of these emphasize not the dispositional contents but the synthesizing and integrating properties of personality, construed in terms such as ego functions, developmental levels, or coping and adaptational mechanisms (Haan, 1977). In contrast to trait models, these approaches focus more on moment-to-moment processes by which experiences are interpreted and actions are guided. Changes in these functions or processes may or may not occur with age, but the present results suggest that the structure or organization of personality dispositions is unaffected by any such hypothesized changes. Results confirming this conclusion with a somewhat different method have been reported by Monge (1975), who found no difference in the structure of semantic differential ratings of the self-concept across the life span.

The accumulated body of evidence on stability of personality should not mean the end of developmental interest in adult personality. The present data came from a highly educated, male sample, and generalizability of the results to less select populations or to women has yet to be demonstrated. If stability in personality were shown to be characteristic of all adults, developmentalists could profitably turn to the problem of identifying the mechanisms by which dispositions are maintained despite the biological, cognitive, and social changes that aging indisputably brings. A succession of phenotypically different behaviors may be needed at different ages to express the same enduring disposition. The longitudinal study of these behavioral changes seems to fit various suggestions for developmental research (McCall, 1977).

From a practical point of view, gerontologists who employ personality measures should be encouraged by evidence that standard personality tests maintain the important psychometric property of factorial validity when used in older populations. As Lawton et al. (1980) point out, elderly subjects present special problems in the administration of psychological tests because of cognitive, sensory, and sometimes motivational deficits. When such problems (if present) can be overcome by careful instructions and special procedures, the present data suggest that objective tests can yield meaningful data even for subjects of advanced age.

REFERENCES

Baltes, P. B., & Nesselroade, J. R. Cultural change and adolescent personality development. *Developmental Psychology*, 1972, *7*, 244-256.

Bendig, A. W. Age differences in the interscale factor structure of the Guilford-Zimmerman temperament survey. *Journal of Consulting Psychology*, 1960, *24*, 134-138.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. *The* *handbook for the Sixteen Personality Factor Questionnaire.* Inst. for Personality and Ability Testing, Champaign, IL, 1970.

Costa, P. T., Jr., & McCrae, R. R. Age differences in personality structure: A cluster analytic approach. *Journal of Gerontology*, 1976, *31*, 564-570.

Costa, P. T. Jr., & McCrae, R. R. Age differences in personality structure revisited: Studies in validity, stability, and change. *Aging and Human Development*, 1977, *8*, 261-275.

Costa, P. T., Jr., & McCrae, R. R. Objective personality assessment. In M. Storandt, I. C. Siegler & M. F. Elias (Eds.), *The clinical psychology of aging*. Plenum Press, New York, 1978.

Costa, P. T., Jr., & McCrae, R. R. Still stable after all these years: Personality as a key to some issues in aging. In P. M. Baltes & O. G. Brim, Jr. (Eds.), *Life-span development and behavior*. (Vol. III). Academic Press, New York, 1980.

Costa, P. T., Jr., McCrae, R. R., & Arenberg, D. Enduring dispositions in adult males. *Journal of Personality and Social Psychology*, 1980, *38*, 793-800.

Douglas, K., & Arenberg, D. Age changes, cohort differences, and cultural change on the Guilford-Zimmerman temperament survey. *Journal of Gerontology*, 1978, *33*, 737-747.

Eysenck, H. J. *The structure of human personality*. Methuen, London, 1960.

Gergen, K. J. Stability, change, and chance in understanding human development. In N. Datan & H. W. Reese (Eds.), *Life-span developmental psychology: Dialectical perspectives on experimental research*. Academic Press, New York, 1977.

Gorsuch, R. L. *Factor analysis*. W. B. Saunders, Philadelphia, 1974.

Guilford, J. P., & Zimmerman, W. S. *The Guilford-Zimmerman temperament survey: Manual of instructions and interpretations.* Sheridan Supply Co., Beverly Hills, CA, 1949.

Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. The Guilford-Zimmerman temperament survey handbook: Twenty-five years of research and application. Robert R. Knapp, San Diego, 1976.

Haan, N. *Coping and defending*. Academic Press, New York, 1977.

Lawton, M. P., Whelihan, W. M., & Belsky, J. K. Personality tests and their uses with older adults. In J. Birren (Ed.), *Handbook of mental health and aging*. Prentice Hall, New York, 1980. (in press)

Maddi, S. R. *Personality theories: A comparative analysis* (2nd ed.). Dorsey Press, Homewood, IL, 1976.

McCall, R. B. Challenges to a science of developmental psychology. *Child Development*, 1977, *48*, 333-344.

Mitchell, J. V., Jr., & Pierce-Jones, J. A factor analysis of Gough's California Psychological Inventory. *Journal of Consulting Psychology*, 1960, *24*, 453-456.

Monge, R. H. Structure of the self-concept from adolescence through old age. *Experimental Aging Research*, 1975, *1*, 281-291.

Neugarten, B. L. Summary and implications. In B. L. Neugarten (Ed.), *Personality in middle and late life*. Atherton Press, New York, 1964.

Neugarten, B. L. Personality and aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. Von Nostrand Reinhold, New York, 1977.

Riley, M. W. Social gerontology and the age stratification of society. *Gerontologist*, 1971, *11*, 79-87.

Schaie, K. W. A general model for the study of developmental problems. *Psychological Bulletin*, 1965, *64*, 92-107.

Siegler, I. C., George, L. K., & Okun, M. A. Cross-sequential analysis of personality. *Developmental Psychology*, 1979, *15*, 350-351.

Cander, L., and Moyer, J.H. (Eds.),
Aging of the Lung: Perspectives
New York: Grune and Stratton, 1964

Changes in Ventilation with Age

By ARTHUR NORRIS, B.A., CHARLES MITTMAN, M.D. AND
NATHAN W. SHOCK, PH.D.

THE LARGE VARIATION in estimates of pulmonary ventilation may be accounted for by voluntary alteration in breathing pattern, involuntary alteration secondary to emotional changes, age, sex, state of health and environmental factors such as altitude, humidity and smog. In addition, it is conceivable that factors of selection or the lessened physical activity of institutional residents accounts for lower functional competence.

The well-known age differences in lung volumes[1,2] show that when body size is taken into account, the total lung capacity is not affected by age itself. In old age, vital capacity (VC) is reduced and there is a complementary increase in the residual volume (RV) of the lung. Maximum breathing capacity (MBC) is lower in older people. When the lung volumes of a sample of active people residing at home were compared with a group of residents of the Baltimore City Hospitals Old People's Home (after the elimination of individuals with clinical signs of impairment of ventilatory function from both groups), significant differences in VC, MBC and RV remained between the groups.[3]

Little attention is paid to ventilatory function unless it limits daily activity or fails to maintain homeostasis. Even in the aged, the ventilation is adequate to meet the metabolic demands of all ordinary activities.

Shock and Yiengst[4] have shown that measurements of ventilation made under basal conditions do not differ between young and old subjects. We have noted that the ventilatory response to a standard exercise test was greater during and following the exercise and persisted for a longer period of time in the post-exercise period in older people.[5,6] This was true for wide ranges of submaximal exercise levels and ages. Thus, the ventilation associated with a given amount of work was always greater in the older group than in the young.[5,7] These findings also pertained if ventilation was related to oxygen uptake instead of work done.

Both Robinson in the U. S.[8] and Valentin and his associates in Germany[9] have measured the maximum ventilatory response and maximum oxygen uptake produced by exercise in young and old persons. The levels of work performed by the older subjects were lower than those performed by the young. The maximum oxygen uptake in the old (1.6 L./min.) were 50 per cent of those found in the young (3.2 L./min.) while the maximum ventilation found in the old (75 L./min.) were 60 per cent of those found in the young

---

From the Gerontology Branch, National Heart Institute, National Institutes of Health, PHS, U. S. Department of Health, Education and Welfare, Bethesda and Baltimore City Hospitals, Baltimore, Md.

(120 L./min.). Thus, even under these conditions older persons had higher ratios of ventilation to oxygen consumed than did the young.

One possible explanation for this decreased ventilatory efficiency is that the distribution of inspired air is more uneven in older people. Bouhuys[10] has concluded that aging alone does not account for the observed impairment of pulmonary gas distribution in elderly people. He suggested that the impairment could be related to long-term effects of inhaled noxious agents.

It is our purpose here to re-examine this finding in a group of healthy people who live in their own homes and to compare the ventilatory responses of this group to those found in a group of residents of similar age in the Baltimore City Hospitals Old People's Home. In addition, we will present the results of serial measurements of ventilatory function made in the people residing at home to assess age changes.

METHODS

Technics

Spirograms were measured with a 13.5 L. Collins spirometer and maximum breathing capacities (MBC) with a 120 L. Collins spirometer.

Ventilation volumes were measured either with a 120 L. Collins spirometer and a stopwatch, or with a pair of 10 L. Krogh spirometers which filled and emptied alternately and sequentially.[7]

Breathing rates were recorded from a pressure switch in the expired air circuit or counted with a stopwatch during visual monitoring.

Nitrogen concentration curves were recorded during oxygen breathing with a Med-Science nitrogen analyzer which sampled expired gas at the mouthpiece.

Subjects

Subjects were drawn from a group of 500 participants in the longitudinal studies of aging in progress in our laboratories. All participants live in their own homes and are members of an active, successful segment of the population of the Washington-Baltimore area. The average level of education in the group is remarkably high (college plus 1 year of graduate training). This self-recruiting group ranges in age from 20 through 102 years. Subjects are tested in our laboratories for a 3-day period every 18 months. Expired minute volume (V_E), breathing rate (f), vital capacity (VC) and maximum breathing capacity (MBC) measurements made during the first 3-day series of tests are reported for 408 members of this group. Repeated measurements of MBC are reported for 109 members of the group and initial analyses of nitrogen washout curves for 47 members of the group.

Subject selection was practiced only for the evaluation of the nitrogen washout curves. Three participants were eliminated from the group of 50 who were tested. Two individuals had obstructive pulmonary disease and another claimed a cigarette smoking rate of 5-6 packs per day.

Calculations

Analyses of the nitrogen washout curves were essentially those described by Fowler, Cornish and Kety.[11] However, as suggested by Lundin,[12] breath-

Table 1.—*Mean Values and Standard Deviations of the Distribution of Individual Values for Minute Ventilation, Breathing Rate, Vital Capacity and Maximum Breathing Capacity* by Age Decade Groups. Values for Minute Ventilation Are Reduced to STPD; Vital Capacity and Maximum Breathing Capacity Are Expressed in BTPS*

Age Group	Minute Volume (L/min.)	Breathing Rate (breadths/min.)	Vital Capacity (L.)	Breathing Capacity (L./min.)
20–29 yr.	5.92	13.1	5.65	165.7
n = 13	± .83	± .8	± .84	± 32.6
30–39 yr.	6.21	14.7	4.96	146.7
n = 70	± 1.00	± 2.5	± 1.18	± 29.4
40–49 yr.	6.31	14.6	4.82	139.1
n = 100	± 1.09	± 2.8	± .77	± 30.0
50–59 yr.	6.21	14.6	4.49	127.8
n = 93	± .88	± 2.5	± .70	± 26.2
60–69 yr.	6.27	14.2	4.18	117.3
n = 71	± 1.25	± 1.2	± .79	± 27.3
70–79 yr.	6.22	14.4	3.87	106.2
n = 55	± .89	± 2.3	± .67	± 30.9
80–89 yr.	5.91	13.2	3.38	75.5
n = 8	± 2.63	± .6	± .68	± 32.1

*This table includes data from first visit tests only.

by-breath end tidal nitrogen concentration rather than mean expired nitrogen concentration was measured. The logarithm of nitrogen concentration for each breath was plotted against number of breaths for each washout curve. Curves were analyzed in terms of single or double exponential slopes. Standard methods were used to compute various indices from the curves obtained. These indices included the size and N_2 washout rates of the compartments, the apparent alveolar ventilation from slope analysis and the degree of uneven ventilation as expressed by the pulmonary clearance delay percentage.[11]

The lung clearance index (LCI) was calculated as the ratio of the total amount of oxygen ventilated to reach a 2 per cent nitrogen concentration in the expired air to the FRC.

Alveolar ventilation was calculated for age decade groups as follows: $\dot{V}_A = \dot{V}_E - \text{D. S. (wt.)} \times f$, where $\dot{V}_A$ = alveolar ventilation, and D. S. (wt.) = an approximation of dead space from body weight in pounds. The dead space approximation was suggested by Radford.[13]

Calculations of mean values, standard deviations and linear regression coefficients were performed according to methods described by Snedecor.[14]

RESULTS

Table I shows the mean values and standard deviations for $\dot{V}_E$, f, VC and MBC for first-visit tests in 408 subjects. Differences between young and old persons were found for VC and MBC but not for $\dot{V}_E$ and f. Regressions of VC and MBC on age were 0.035 L. per year and 1.244 L./min. per year re-

Table 2.—*Calculated Alveolar Ventilations and Observed Minute Volumes* Are Compared by Age Decades for the Community Residing Participants in the Longitudinal Studies (A) and a Group of Residents of the Baltimore City Hospitals Old People's Home (B). Volumes Are Reduced to L./min. STPD*

Age Group	No. of Subjects		Minute Volume		Alveolar Ventilation	
	A	B	A	B	A	B
20–29 yr.	13		5.92		3.64	
30–39 yr.	72		6.21		3.[illegible]7	
40–49 yr.	99	27	6.31	6.90	3.80	4.54
50–59 yr.	93	27	6.21	6.95	3.67	4.64
60–69 yr.	69	39	6.27	6.71	3.80	4.34
70–79 yr.	54	38	6.22	6.87	3.89	4.37
80–89 yr.	8	21	5.91	6.57	3.87	4.19

*This table includes data for first visit tests only.

Table 3.—*Mean Values by Age Decade Groups of Maximum Breathing Capacity for Subjects Who Had Two or Three Measurements Made on Separate Visits 18 Months Apart. MBC Measurements Are Expressed as L./min. BTPS*

Age Group	Visit No.			No. of Subjects
	1	2	3	
20–29 yr.	177	196	—	6
30–39 yr.	139	161	160	10
40–49 yr.	130	150	147	20
50–59 yr.	124	135	140	24
60–69 yr.	119	127	132	23
70–79 yr.	106	113	105	21
80–89 yr.	70	85	90	5

spectively. The levels of regression curves at age 55 years were 4.53 L. for VC and 129.1 L./min. for MBC.

Table 2 shows that when the present group of subjects (Group A) was compared to the residents of the Old People's Home (Group B), there were no age differences for $\dot{V}_E$ or calculated $\dot{V}_A$ within either group and that observed $\dot{V}_E$ and calculated $\dot{V}_A$ were significantly ($p = <.001$) higher for Group B than for Group A.

When changes with time were sought by comparing measurements of $\dot{V}_E$, f, VC and MBC, which were made on three different occasions (18 months apart) in the same group of individuals, significant ($p = <.001$) differences were found *only* between the first and second visits for MBC (shown in table 3). Group averages for the other measures were found to be quite similar from visit to visit.

Nitrogen washout curves were analyzed for 47 individuals in Group A. Single exponential curves corresponding to uniform lung ventilation were found for four of these individuals. Significant ($p = <.001$) age differences were found only for LCI. Table 4 gives the mean values, range of values and standard deviation of the distribution for five age groups. The slope of re-

Table 4.—*Mean Values, Range of Values and Standard Deviations of the Distributions of Lung Clearance Index (LCI) by Age Groups*

Age Group	Mean Value	Range of Values	Standard Deviation	Number of Subjects
29–39 yr.	9.17	7.91–11.04	1.02	7
40–49 yr.	13.66	10.82–19.80	3.20	7
50–59 yr.	12.06	9.62–14.80	2.02	11
60–69 yr.	12.87	9.96–18.00	2.20	12
70–90 yr.	14.86	11.28–17.36	2.05	10

Table 5.—*Comparison of Mean Values and Standard Errors of the Mean Values of Functional Residual Capacity (FRC) between Data of Bouhuys (1963) and the Present Study*

Bouhuys			Present Study		
Age Group	No. of Subjects	Mean and Standard Error	Age Group	No. of Subjects	Mean and Standard Error
24–34 yr.	19	2.67 ± 0.14	29–39 yr.	7	3.12 ± 0.14
35–44 yr.	24	2.63 ± 0.13	40–49 yr.	7	2.62 ± 0.30
45–54 yr.	22	2.93 ± 0.12	50–59 yr.	11	2.44 ± 0.17
55–65 yr.	15	2.95 ± 0.17	60–69 yr.	12	2.69 ± 0.21
			70–90 yr.	10	2.29 ± 0.16

gression of LCI on age for the entire group was 0.107 units per year. The level of the regression curve was 12.42 LCI units at age 55 years.

DISCUSSION

It is important to distinguish between age differences and age changes. Differences between two or more systems and changes within a system should not be confused. We may not infer from the existence of a significant difference between a sample of young survivors and a sample of aged survivors that the young group will exhibit changes of a direction and degree which will make them like the aged group when they have survived a similar length of time. Changes may be followed by making repeated measurements in an individual or group. Two relationships between age differences and age changes are demonstrated by the results reported here. In the case of $\dot{V}_E$ and f, neither age differences nor age changes were observed. In the case of VC and MBC, significant age differences were found while age changes were not demonstrated over a 36-month interval.

As these measurements are continued over a longer period of time, changes may appear. Possibly, after a period of stability of the function measured, an abrupt change may occur. If such change can be detected in an individual in spite of a variability of the measurement, it may be related to an accident, disease incident, or another change in function. On the other hand, slower systematic changes may not be seen in a 36-month interval but may emerge after many years of testing. Slower changes are consistent with the net loss of a few cells each day or month or year from the organs concerned. In addition, a combination of abrupt and slower changes may be found.

The difference in basal ventilation between the group reported here and the group of residents of the Baltimore City Hospitals Old People's Home[4] emphasizes the need to make repeated observations in individuals and groups rather than to characterize age differences by comparing young persons with old persons whose history cannot be determined with precision.

The MBC is a performance test which requires the intelligent cooperation of the subject. The appropriate rate and volume for optimal performance by an individual can be learned. We have assumed that this learning could be accomplished during the explanation and three trials which were performed on the first visit to the laboratory. However, the marked increase in average performance between first and second visits casts doubt on this assumption. We have no reason to believe that the participants practice the MBC test between visits or just prior to a visit 18 months later and we prefer to suggest that greater familiarity with the procedure and a more relaxed performance result in higher scores on visits following the first.

The significant age differences in lung clearance index reported here agrees with the results reported by Bouhuys.[10] His average values, however, ranged from 9.0 ± 0.2 in his youngest subjects (24-34 years old) to 10.0 ± 0.6 in his oldest subjects (55-65 years old), and were lower than the values reported here. This was a result of the higher values for functional residual capacity (FRC) found by Bouhuys in his older subjects than were found in the older subjects of the present study (table 5). Bouhuys' values for FRC were as high or higher than those we have reported for residents of the Baltimore City Hospitals Old People's Home.[2]

It has been suggested that reduced compliance of the lung is an underlying factor in pulmonary changes found in advanced age.[15,16] The larger ventilatory response to submaximal exercise[5,7] which occurs in older people, and the higher ratios of ventilation to oxygen consumed found in older persons when maximum oxygen uptake is induced by exercise[8,9] may be related to age differences in ventilation reported here as well as to limitations of diffusing capacity for oxygen[17] and limitation of blood flow[18] which have been found in older people.

SUMMARY

Minute volume and breathing rate were found to be similar for young and old persons in a group of 408 participants in a longitudinal study of aging. Vital capacity and maximum breathing capacity decreased in the older participants. When 108 of the participants were tested a second and a third time, 18 months and 36 months after the first visit, no changes were found for any age group excepting an increase in maximum breathing capacity between the first and second visits in all age groups. Thus, age differences between young and old individuals or groups of individuals do not necessarily predict changes in the population sampled and certainly should not be used to anticipate changes in other populations. There is a significant delay of nitrogen washout in older subjects as compared with younger subjects in a sub-group of 47 individuals from the population reported here.

ACKNOWLEDGMENTS

We wish to thank John B. Melvin, Earl Crizer, Kristin Steuber, and Lorraine Ward who performed the tests and calculations and Dr. Norman Edelman who assisted with the recording and analysis of the nitrogen washout curves.

REFERENCES

1. Baldwin, E. deF., Cournand, A., and Richards, D. W., Jr.: Pulmonary insufficiency: I. Methods of analysis, physiological classification, standard values in normal subjects. Medicine 1948, 27:243.
2. Norris, A. H., Shock, N. W., Landowne, M., and Falzone, J. A., Jr.: Pulmonary function studies; age differences in lung volumes and bellows function. J. Gerontol. 1956, 11:379.
3. —, —, and Falzone, J. A., Jr.: Relation of lung volumes and maximal breathing capacity to age and socio-economic status. *In* Medical and Clinical Aspects of Aging. H. T. Blumenthal, ed. New York, Columbia Univ. Press, 1962, p. 162.
4. Shock, N. W., and Yiengst, M. J.: Age changes in basal respiratory measurements and metabolism in males. J. Gerontol. 1955, 10:31.
5. Norris, A. H., and Shock, N. W.: Age changes in ventilatory and metabolic responses to submaximal exercise. *In* Fourth Congress of International Association of Gerontology, Vol. 2. Fidenza, Tito Mattioli, Fidenza, 1957, p. 512.
6. —, Falzone, J. A., Jr., and Shock, N. W.: Age differences in the time curve of recovery of metabolic processes following graded exercise. Fed. Proc. 1953, 12:105.
7. —, —, and Yiengst, M. J.: Age differences in ventilatory and gas exchange responses to graded exercise in males. J. Gerontol. 1955, 10:145.
8. Robinson, S.: Experimental studies of physical fitness in relation to age. Int. Z. Angew. Physiol. 1938, 10:251.
9. Valentin, H., Venrath, H., von Mallinckrodt, H., and Gurakar, M.: Die maximale Sauerstoffaufnahme in den verschiedenen Altersklassen. Eine praktisch wichtige Herz-kreislauf-funktionsprüfung im Vita-maximabereich. Ztschr. Altersforsch. 1955, 9:291.
10. Bouhuys, A.: Pulmonary nitrogen clearance in relation to age in healthy males. J. Appl. Physiol. 1963, 18:297.
11. Fowler, W. S., Cornish, E. R., Jr., and Kety, S. S.: Lung function studies. VIII. Analysis of alveolar ventilation by pulmonary N_2 clearance curves. J. Clin. Invest. 1952, 31:40.
12. Lundin, G.: Alveolar ventilation in normal subjects analyzed breath by breath as nitrogen elimination during oxygen breathing. Scandinav. J. Clin. & Lab. Invest. 1955, 7: suppl. 20, 39.
13. Radford, E. P., Jr.: Ventilation standards for use in artificial respiration. J. Appl. Physiol. 1955, 7:451.
14. Snedecor, G. W.: Statistical Methods, 5th ed. Ames, Iowa State College Press, 1956.
15. Cohn, J. E.: Mechanical properties of human lungs in normal males over 60 years of age. *In* Sixth International Congress of Gerontology, International Congress Series No. 57. Amsterdam, Exceptra Medica Foundation, 1963.
16. —, and Donoso, H. D.: Mechanical properties of lung in normal men over 60 years old. J. Clin. Invest. 1963, 42:1406.
17. —, Carroll, D. G., Armstrong, B. W., Shepard, R. H., and Riley, R. L.: Maximal diffusing capacity of the lung in normal male subjects of different ages. J. Appl. Physiol. 1954, 6:588.
18. Brandfonbrener, M., Landowne, M., and Shock, N. W.: Changes in cardiac output with age. Circulation 1955, 12:557.

American Journal of Epidemiology
Vol. 110, No. 6

OSTEOARTHRITIS OF THE HAND: LONGITUDINAL STUDIES[1]

CHRIS C. PLATO AND ARTHUR H. NORRIS

Plato, C. C., and A. H. Norris (Gerontology Research Center, National Institute on Aging, Baltimore City Hospitals, Baltimore, MD 21224). Osteoarthritis of the hand: Longitudinal studies. *Am J Epidemiol* 110:740–746, 1979.

Evaluation of the osteoarthritic grades of the hands of 478 participants of the ongoing Baltimore Longitudinal Study suggests that: 1) Joint degeneration due to osteoarthritis is a relatively slow process. The maximum rate of degeneration is seen in the distal interphalangeal joints where the average increase is about 1 grade per individual in an interval of 12 to 16 years between visits in each age group. The rate of degeneration in the proximal interphalangeal joints is much lower than that of the distal interphalangeal joints. 2) The progress of the degeneration in the distal interphalangeal joints of an individual (longitudinally evaluated) follows closely that which is observed at the population level (cross-sectional joint-digit study). That is, it is directly related to the age and the interval between visits. This is not always seen in the proximal interphalangeal joint data. 3) The rate of change in the osteoarthritic grade of individual hands agrees closely with that of their distal interphalangeal joints. This further supports the conclusions reached in a first report that what has been referred to as osteoarthritic grade of the hand of an individual may actually be the higher grade among the distal interphalangeal joints.

aged; hand; joint diseases; joints; longitudinal studies; osteoarthritis; osteoarthropathy

The objective of this ongoing epidemiologic investigation of osteoarthritis of the hand is to determine the rate of change in joint degeneration within different age groups.

MATERIALS AND METHODS

The radiographs utilized in the longitudinal study were of the left hand of participants in the Baltimore Longitudinal Study of the Gerontology Research Center (GRC) of the National Institute on Aging. All participants were male Caucasians between 21 and 97 years of age. They were generally in good health, with above average education and socioeconomic position. Over 85 per cent were college graduates in professional, technical and

---

Received for publication January 22, 1979, and in final form April 2, 1979.

Abbreviations: A, B, C, and D age groups—<40, 40–54, 55–69, and ≥70 years, respectively; DH, IH, and PH—highest osteoarthritic grade encountered in, respectively, any of the 5 distal interphalangeal joints of a given hand, any of the digital joints of a given hand, and any of the 5 proximal interphalangeal joints of a given hand. PSI, prevalence severity index. V_0, initial visit; V_1, V_2, V_3, V_4—return visit 0–3, 4–7, 8–11, 12–16 years, respectively, after V_0. +, affected or positive, osteoarthritic grades 2, 3, or 4; ++, deterioration from positive to more severe form. −, non-affected or negative, osteoarthritic grades 0 or 1.

[1] From the Human Performance Section, Gerontology Research Center, National Institute on Aging, National Institutes of Health, Baltimore City Hospitals, Baltimore, MD 21224 (reprint requests to Dr. Plato).

The authors are indebted to Drs. Stanley M. Garn, Ian Higgins, William M. Mikkelsen and Andrew Poznanski, all from The University of Michigan, for their invaluable suggestions during the course of this investigation.

managerial positions. Details of the GRC sample and the method of its ascertainment were described in the first article of the series (1) as well as in a report by Stone and Norris (2).

The longitudinal study of osteoarthritis of the hand involved 478 participants who had been radiographed at two or more visits. Since not all participants had the same number of x-rays and the interval between successive x-rays was not always the same, certain adjustments had to be made. The visit at which the first x-ray was taken, V_0, formed the basis of these adjustments. All subsequent radiographs of the same individual were grouped depending on the number of years that elapsed since the first visit. Group V_1 included all x-rays taken within three (0–3) years after V_0, with a mean interval of 2.28 years. Group V_2 included the x-rays of each participant taken within 4–7 years after his first x-ray, with a mean of 5.85 years. Group V_3 included x-rays repeated within 8–11 years after V_0, with a mean of 9.47 years. Radiographs in group V_4 were taken 12 to 16 years after the initial visit, with a mean interval of 13.45 years.

The participants were grouped into four age groups according to their age at V_0. Group A included 107 individuals below the age of 40, with a mean age of 33.2 years; group B was composed of 175 participants age 40 to 54, with a mean age of 46.9 years; group C included 129 participants age 55 to 69, with a mean age of 61.8 years; and group D consisted of 67 individuals who were x-rayed for the first time at the age of 70 years or older, with a mean age of 74.1 years.

The x-rays were graded following the standards suggested by Kellgren (3) which are also described in our previous report (1). Under this classification there are five osteoarthritic grades or scores. These grades, listed in ascending order of severity, are:

Grade	Degree of involvement
0	None
1	Doubtful
2	Minimal
3	Moderate
4	Severe

Grades 0 and 1 are considered normal (−) while joints scored 2, 3 or 4 are considered affected (+).

The data on osteoarthritis in the distal interphalangeal and proximal interphalangeal joints were recorded and evaluated separately using the DH, PH, and IH scores. The DH scores represent the highest osteoarthritic grade among the five distal interphalangeal joints and the PH the highest grade among the four proximal interphalangeal joints plus the metacarpophalangeal joint of the thumb (1). The IH is the highest osteoarthritic grade observed in any of the interphalangeal or metacarpophalangeal joints of the hand. The IH corresponds to the osteoarthritic grade of the hand reported in the literature (3). The inclusion of joints whose score was two or more grades higher than the rest of the joints may have resulted in an increase in the reporting of traumatic osteoarthritis cases.

Since osteoarthritis in the metacarpophalangeal joints is much less prevalent and not as severe as in the interphalangeal joints, it was not included in the present analysis.

All radiographs were evaluated by the same investigator, which eliminates inter-investigator variability. Kellgren and Lawrence (4) reported 0.73 and 0.81 inter- and intra-observer correlation, respectively, for the distal interphalangeal joints. In this study, the intra-investigator variability was not calculated. There were, however, 25 retrogressive gradings (1.3 per cent of the 1989 longitudinal evaluations). Seventeen of these involved discrepancies between grades 1 and 0, seven between grades 2 and 1 and two between grades 4 and 3. The final grade in these cases was determined by re-examining the x-rays of both visits.

Results and discussion

The longitudinal increases in the grades of DH, PH, and IH are presented in tables 1 and 2. These tables show the age of the participants at the initial visit (V_0) and indicate the osteoarthritic status of the same set of joints at V_0 and at any of the subsequent visits (V_i). When the data under the heading "At V_0 → At V_i" are taken together, code – → + indicates an individual whose hand was found normal (grades 0 or 1) at V_0, but had become affected at any of the subsequent visits (V_i). Likewise, the code + → ++ indicates a participant whose x-ray was graded 2 or 3 (affected) at V_0 and deteriorated further (was given a higher grade) at V_i. The $V_0 < V_i$ code (the grade at V_0 was lower than that at V_i) includes all participants whose hand x-rays at any of the subsequent visits were given higher osteoarthritic grades than those of their respective first visits; this includes changes from grade 0 to grade 1 even though both are considered to be normal. The discussion of the longitudinal study will be concerned mainly with the $V_0 < V_i$ changes.

The values in tables 1 and 2 shown under "No. at V_0," give the number of individuals at V_0 whose grade changed later as indicated under "Osteoarthritic status." Values under the headings "% changed at V_1" and at "V_2," "V_3," "V_4," indicate the proportion of individuals whose osteoarthritis status changed during the interval from V_0 to V_1, from V_0 to V_2, from V_0 to V_3, and from V_0 to V_4. In other words, the grades at V_1, V_2, V_3, and V_4 are compared to the grades of V_0 and not to the grades of each other. It should be noted here that participants whose x-rays were given a grade 4 at V_0 were excluded from the longitudinal study since their grade could not change. The $V_0 < V_i$ values of tables 1 and 2 are also presented in a series of graphs as shown in figure 1.

In figure 1 the abscissa indicates the four age groups, <40 (A), 40–54 (B), 55–69 (C) and 70+ years of age (D). The readings along the solid (vertical) lines indicate the per cent of individuals, within each age group, whose osteoarthritis grade was found to be higher (deteriorated) in visits V_1, V_2, V_3 or V_4 than what it was at the initial visit, V_0. The readings along the broken lines indicate, within each visit, the age distribution of participants whose osteoarthritic grade deteriorated from that of V_0.

Between visits-within age groups. The between visit-within age group changes are distributed along the solid vertical lines of figure 1. They show that, in all joints, the longer the interval from the first examination (V_0) the higher will be the per cent of individuals whose osteoarthritic condition became worse. In the distal interphalangeal joints, DH, the rate of change in different visit intervals is more or less constant within each age group. The main exception is seen in the wider gap between the V_2 and V_3 among the participants of the 70+ age group. In the proximal interphalangeal joints, PH, the per cent of individuals showing a higher grade than at V_0 varies from visit to visit within each age group. The IH (highest osteoarthritic grade observed in the interphalangeal or metacarpophalangeal joints) within age distributions resembles those of the distal interphalangeal joints.

Between age groups-within visits. The between age group-within visit changes are indicated along the broken lines in figure 1. In the distal interphalangeal joints, the rate of worsening of osteoarthritis within each visit interval (along the broken lines) increases steadily with age. The only exceptions to this trend are the slight decreases in the percentages of V_1 and V_2 in the 70+ age group. These exceptions are probably the result of a rarity

TABLE 1

Per cent changes in highest osteoarthritic status in distal (DH) interphalangeal joints from initial visit to re-examination at any subsequent visit (V_i): 0–3 years (V_0/V_1), 4–7 years (V_0/V_2), 8–11 years (V_0/V_3), and 12–16 years later (V_0/V_4)

Code and years of age at V_0	Osteoarthritic status‡	DH							
		V_0/V_1		V_0/V_2		V_0/V_3		V_0/V_4	
	At V_0 → At V_i	No. at V_0	→ %† changed at V_1	No. at V_0	→ %† changed at V_2	No. at V_0	→ %† changed at V_3	No. at V_0	→ %† changed at V_4
	→ +	16	0	65	7.7	26	23.1	19	36.8
A: 40	+ → + +	1	0	4	0	3	33.3	1	0
	All V_0 – V_i*	17	5.9	69	27.5	29	44.8	20	60.0
	→ +	12	8.3	97	24.7	27	22.2	36	44.4
B: 40–54	+ → + +	4	25.0	41	29.3	18	50.0	11	63.6
	All V_0 – V_i*	16	25.0	138	39.1	45	53.3	47	70.2
	→ +	2	0	46	39.1	20	50.0	9	77.8
C: 55–69	+ → + +	4	25.0	57	38.6	31	61.3	14	78.6
	All V_0 – V_i*	6	33.3	103	42.7	51	62.7	23	78.3
	→ +	3	66.7	12	66.7	4	100.0	1	100.0
D: 70+	+ → + +	13	15.4	34	26.5	8	75.0	3	100.0
	All V_0 – V_i*	16	25.0	46	37.0	12	83.3	4	100.0
	→ +	33	9.1	220	25.0	77	33.8	65	47.7
E: All	+ → + +	22	18.2	136	31.6	60	58.3	29	72.4
	All V_0 – V_i*	55	20.0	356	37.6	137	57.7	94	71.3

* Does not include grade 4 at V_0.

† Includes changes from 0 to 1.

‡ +, positive; + → + +, deterioration from positive to more severe form.

TABLE 2

Per cent changes in highest osteoarthritic status in proximal (PH) interphalangeal joints from initial visit to re-examination at any subsequent visit (V_i): 0–3 years (V_0:V_1), 4–7 years (V_0:V_2), 8–11 years (V_0:V_3), and 12–16 years later (V_0:V_4)

Code and years of age at V_0	Osteoarthritic status‡	PH							
		V_0:V_1		V_0:V_2		V_0:V_3		V_0:V_4	
	At V_0 → At V_i	No. at V_0 →	% changed at V_1	No. at V_0 →	% changed at V_2	No. at V_0 →	% changed at V_3	No. at V_0 →	% changed at V_4
	− → +	17	0	69	4.4	28	10.7	19	21.1
A: <40	+* → + +	0	0	0	0	1	0	1	0
	All* (V_0 → V_i)†	17	11.8	69	21.7	29	44.8	20	65.0
	− → +	13	15.4	126	14.3	41	22.0	43	32.6
B: 40–54	+* → + +	3	0	13	7.7	5	20.0	4	0
	All* (V_0 → V_i)†	16	25.0	139	28.1	46	47.8	47	53.2
	− → +	3	0	81	23.5	40	42.5	15	46.7
C: 55–69	+* → + +	4	25.0	25	20.0	14	14.3	8	37.5
	All* (V_0 → V_i)†	7	28.6	106	37.7	54	50.0	23	56.5
	− → +	7	0	28	32.1	7	42.9	4	75.0
D: 70+	+* → + +	8	12.5	19	31.6	6	50.0	1	0
	All* (V_0 → V_i)†	15	13.3	47	38.3	13	46.2	5	60.0
	− → +	40	5.0	304	16.1	116	27.6	81	34.6
E: All	+* → + +	15	13.3	57	21.1	26	23.1	14	21.4
	All* (V_0 → V_i)†	55	18.2	361	31.0	142	47.9	95	56.8

* Does not include grade 4 at V_0.

† Includes changes from 0 to 1.

‡ +, positive; + → + +, deterioration from positive to more severe form.

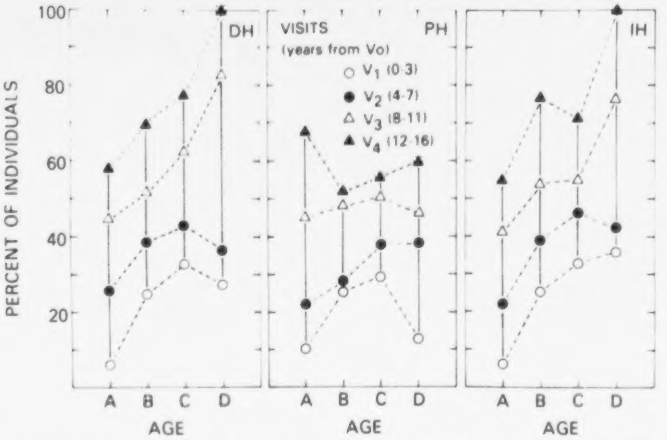


FIGURE 1. Frequency (per cent) of participants whose osteoarthritic grade in distal (DH), proximal (PH), and all hand joints (IH) was higher in subsequent visits (V_1, V_2, V_3, or V_4) than that of their first visit (V_0) for age group < 40 years (A), 40 – 54 (B), 55 – 69 (C), and 70+ (D). Connections with broken lines show differences between age groups within visits. Connections with solid lines show differences between visits within the same age group.

of normal individuals who will eventually become affected (0 → + type, since in the older age group one sees mostly + → ++ changes).

The above exceptions not withstanding, the overall distribution of the DH frequencies suggests that in the distal interphalangeal joints the progression of osteoarthritic disease is positively related to increasing age as well as to increasing visit interval from the initial examination.

The "within" visit distributions of the proximal interphalangeal joints, PH, are not associated with the changing age groups. In these joints the within visit per cent of individuals showing osteoarthritic worsening fluctuates indiscriminately or remains unchanged from one age group to the other. The IH distributions are again similar to those of DH.

In summary, the data of figure 1 suggest that 1) osteoarthritis of the distal interphalangeal joints deteriorates progressively with increasing intervals from the initial examination and 2) the rate of deterioration is enhanced with increasing age. On the other hand, osteoarthritis of the proximal interphalangeal joints 1) deteriorates less predictably with increasing visit interval from V_0 and 2) shows a rate of worsening of the condition which is not related to the age of the individual. The reason for these differences is not clear. It is possible that osteoarthritis in the distal interphalangeal joints is a progressively deteriorating condition due to biological factors associated with aging and is, therefore, unavoidable, while the proximal joint osteoarthritis may be caused by extraneous factors whose effect is increasing in terms of years of exposure rather than biological age. The IH data are included in this figure as added proof of the conclusion of the joint-digit study (1) that what is universally referred to as the osteoarthritic grade of a hand is essentially the DH value.

A series of chi-square tests supported the above observations at the $p < 0.01$ level of significance. First, in all age groups of DH, the longer the interval from V_0 the higher ($p < 0.01$) was the number of individuals whose osteoarthritic condition deteriorated. Second, in the PH comparisons, only in the younger age groups

TABLE 3

Average joint and age specific increase in prevalence-severity index per individual from the first visit (V_0) to each of the other visits

Visits‡ compared from: to	DH*				PH†			
	Age groups§				Age groups§			
	< 40	40–54	55–59	70+	< 40	40–54	55–69	70+
V_0:V_1	0.06	0.25	0.33	0.25	0.12	0.25	0.29	0.13
V_0:V_2	0.32	0.49	0.45	0.37	0.23	0.33	0.42	0.38
V_0:V_3	0.52	0.64	0.69	0.92	0.48	0.59	0.56	0.54
V_0:V_4	0.90	0.98	0.96	1.25	0.80	0.64	0.65	0.60

* Highest osteoarthritic status in distal interphalangeal joints.

† Highest osteoarthritic status in proximal interphalangeal joints.

‡ V_1, V_2, V_3, V_4: return visit 0–3, 4–7, 8–11, 12–16 years, respectively, after first visit (V_0).

§ Age at first visit.

(A and B) was there a significant deterioration of osteoarthritis with the longer visit intervals.

The longitudinal aspect of osteoarthritic changes was also evaluated through the prevalence-severity index, PSI, which was introduced in the earlier study (1). This was done by subtracting the actual osteoarthritic score of V_0 from that of the V_i visit. The results are summarized by age groups in table 3. The row readings in this table, from left to right, relate to specific visit intervals and show the average osteoarthritic grade increase per individual in each age group. The column readings, from top to bottom, relate to individual age groups and give the average osteoarthritic grade change per individual in each visit interval. The results of table 3 agree closely with those of the previous paragraphs, which dealt with gross grade changes. This table also gives a quantitative value to the hypothesis that osteoarthritis of the hands is a slowly progressing disease. The highest rate of joint degeneration is observed in the DH comparisons of V_0:V_4 at ages of 70 years and older (see last row of table 3). In this case the disease progressed at an average rate of 1.25 PSI (grade per individual) within an interval of 12–16 years. This value of PSI decreases with decreasing age group and visit interval. Furthermore, the average rate of degeneration in the distal joints (DH) is almost twice as high as that of the proximal joints (PH). This holds true for all visit intervals. The PH rates of joint degeneration do not seem to be affected by age. This is in contrast to what was seen in the distal joint changes.

REFERENCES

1. Plato CC, Norris AH: Osteoarthritis of the hand: Age specific joint-digit prevalence rates. Am J Epidemiol 109:169–180, 1979
2. Stone JD, Norris AH: Activities and attitudes of participants in the Baltimore longitudinal study. J Gerontol 21:575–580, 1966
3. Kellgren JH: The Epidemiology of Chronic Rheumatism. Vol II. Atlas of Standard Radiographs of Arthritis. Philadelphia, FA Davis Company, 1963
4. Kellgren JH, Lawrence JS: Radiological assessment of rheumatoid osteoarthrosis. Ann Rheum Dis 16:494–502, 1957

J Am Geriatr Soc 30: 404–409, 1982

Effect of Aging on the Cardiothoracic Ratio of Men

JANE F. POTTER, MD, DARIUSH ELAHI, PhD, JORDAN D. TOBIN, MD, and REUBIN ANDRES, MD

The potential usefulness of chest radiographs in the assessment of physical dimensions was examined in 243 men (age range 20–95 years) who had been followed up for an average of 12.3 years. From 1,124 of these films, measurements of cardiac diameter (CD) and thoracic diameter (TD) were made, and the cardiothoracic ratio (CTR) was calculated. Cross-sectional age differences were associated with a tendency toward increases in the CD and CTR throughout the lifespans of the 243 men, regardless of the presence or absence of heart disease. The thoracic diameter (TD) was greater in middle-aged than in young subjects, but less in the oldest than in the middle-aged subjects. Cumulative percentage curves for the CTR showed a shift to higher fiftieth to ninetieth percentile values with age. However, among the subjects free of heart disease, only one (age 95) had a CTR exceeding 50 per cent. Longitudinal analysis data agreed with the cross-sectional data. Forty-nine deceased subjects were matched with living subjects of the same heart disease classifications. Increases in the CD and CTR were predictive of death in the group with heart disease but not in the group without identifiable heart disease. An increase in CD was not correlated with an increase in systolic blood pressure. The decline in TD appeared to reflect a decline in rib-cage mobility with aging.

Two kinds of errors can be made in interpretation of laboratory results obtained for elderly patients: 1) for tests that *do not* show changes with normal aging, a truly abnormal result may be misclassified as normal on the incorrect assumption that normal aging has induced the change; 2) for tests that *do* show changes with normal aging, a normal result may be misclassified as abnormal because of the failure to appreciate that age-adjustment of normative criteria should have been applied. The studies reported in this paper indicate that the cardiothoracic ratio (CTR) of men belongs in the former category, and that age-adjustment of the usual standard of normality is not justified.

The cardiothoracic ratio (CTR) calculated from radiographic data is an important clinical measure for assessing the normality of the heart. The usual standard of normality is a CTR less than 50 per cent, as first suggested by Danzer[1] in a report of a study that does not state the ages of the subjects. Cross-sectional studies of persons more than 65 years of age,[2-4] however, show CTR values exceeding 50 per cent for 10 per cent of older men and 20 per cent of older women. This suggests that the usual standard of normality does not apply after 65 years of age. Although these studies excluded persons who had overt heart disease, rigorous screening for heart disease was not done.

With aging, physiologic changes that might influence CTR occur in both cardiovascular and thoracopulmonary systems. These changes are: 1) increased cardiac left ventricular mass resulting from increased vascular resistance,[5] and 2) decreased chest wall compliance due to decreased rib-cage mobility.[6-8]

The present study has examined the usual standard of normality for CTR in a group of old men who had undergone careful screening for the presence of heart disease. They had been followed with serial radiographic chest examinations that allowed determinations of both cross-sectional differences and longitudinal changes in cardiac diameter (CD), and thoracic diameter (TD), and the CTR.

SUBJECTS AND METHODS

Subjects

All subjects were participants in the Baltimore Longitudinal Study of Aging (BLSA). The characteristics of this population have been reported previously.[9] Subjects spend two and a half days in the Gerontology Research Center at 12-month to 24-month intervals, undergoing a variety of clinical,

---

Received from the Clinical Physiology Branch, Gerontology Research Center, National Institute on Aging, National Institutes of Health, PHS, U.S. Department of Health and Human Services, Bethesda, and Baltimore City Hospitals, Baltimore, Maryland. Presented at the XII International Congress of Gerontology, Hamburg, Federal Republic of Germany, July 1981.

physiologic and psychologic tests. Standard 6-foot posteroanterior chest radiographs were part of the routine testing between 1958 and 1975. They provided the material for this study. Criteria for inclusion in this report were: 1) three or more technically adequate radiographs; 2) eight years or more between first and last radiographic examinations; 3) the patient's age at the time of last radiographic examination was either less than 50 or more than 60 years. All subjects more than 60 years old who met the criteria were studied, and the youngest 60 subjects less than 50 years old were used as a control group.

Of the total of 741 subjects enrolled by 1967 in the BLSA, 243 were selected in this manner. At the time of their last visit, 183 subjects were 60 years of age or older and 60 were less than 50 years old.

Methods

A total of 1,124 radiographs were examined, an average of 4.5 per subject (range, 3–8). The average time span between first and last x-ray examinations was 12.3 years (range, 8–21 years). Measurements of CD and TD were made "blind" with respect to the subject's age and the date of the examination. To measure the CD, a line was drawn horizontally between the clavicles, and a second line (perpendicular to the first) bisected the heart. Measurements were made from the midline (perpendicular line) to the points of greatest deviation of the heart to the right and left, and the sum was taken as the CD. The TD was measured from the internal surfaces of ribs on the right and left sides superior to the costal attachments of the diaphragm at the point where the width of the chest was greatest. Care was taken to measure from the same rib on all radiographs from the same subject and to exclude radiographs that did not show the same depth of inspiration. Measurements were made to the nearest 0.01 cm and a series of randomly selected radiographs were re-examined and observed to be reproducibly measured to 0.1 cm. The cardiothoracic ratio (CTR) was calculated as the ratio of CD to TD.

The medical history, physical examination, and a standard 12-lead electrocardiogram (ECG) were obtained for every subject at each visit. In addition, an exercise test (Master's two-step or treadmill) was performed by each patient when not clinically contraindicated. This evaluation served as the basis for classifying most subjects as to whether they had heart disease (HD+) or did not have heart disease (HD−). Most HD+ subjects had coronary heart disease, a diagnosis that was made when the score was a total of 2 points or more according to the following criteria:

A. Angina pectoris: definite history = 2 points, probable history = 1 point
B. Myocardial infarction: definite clinical evidence = 2 points, probable clinical evidence = 1 point
C. Resting electrocardiogram: definite myocardial infarction = 2 points, probable myocardial infarction = 1 point
D. Post-exercise electrocardiogram (Master's two-step or treadmill): a change from no S-T-J abnormality at rest to S-T-J depression of 1.0 mm or more, and S-T segment horizontal or downward sloping in any leads I, II, aVf, VI–V6 = 1 point

Subjects were also considered to be HD+ if: 1) they were receiving digoxin with or without a diuretic for congestive heart failure or arrhythmias; 2) they had a history of hypertension, physical (S-4 gallop) and ECG signs of hypertensive cardiovascular disease; 3) mitral regurgitation was present; 4) cardiomyopathy was present. Of the subjects more than 60 years old, 89 were classified as HD+ and 94 as HD−, and all subjects less than 50 years old were HD−. The types and distribution of heart disease were: 74 cases of coronary heart disease; seven cases of digoxin treatment for arrhythmias or congestive failure; four cases of hypertensive cardiovascular disease; two cases of mitral regurgitation; and one case of cardiomyopathy.

Subjects also underwent anthropometric measurements at most visits, which included measurements of chest circumference (greatest circumference at mid-inspiration). Blood pressure measurements also were made under basal (early morning, with the subject at rest) and casual (during the physician's examination) conditions at each visit.

Statistical Methods

Each subject was represented only once in each analysis. In cross-sectional analysis, this one value was obtained at the time of the last visit, i.e., at the subject's oldest age. In the longitudinal analyses of changes in chest radiographic measurements (Fig. 1), the rate of change was computed as the slope of the regression line for each subject on the basis of three or more data points. The change in a variable (delta variable) for a subject was obtained by subtraction of the last-visit datum from the first-visit datum. Cumulative percentage frequencies were calculated from the frequency distribution of CTRs and drawn as cumulative percentage curves (Fig. 2). In the mortality analysis, deceased subjects were compared with matched survivors by performing a t-test on the difference in mean slopes between the two groups. Mean slopes were calculated from individual slopes in each group after weighting the slopes for variance.[10] Cross-sectional differences were assessed by the Student's t-test. Simple linear correlations and regressions were used to assess the relationships between variables.

RESULTS

The first part of the analysis concerned data at the time of each subject's last visit and yielded, therefore, a cross-sectional view of the study population at its oldest age. The CDs tended to be greater in

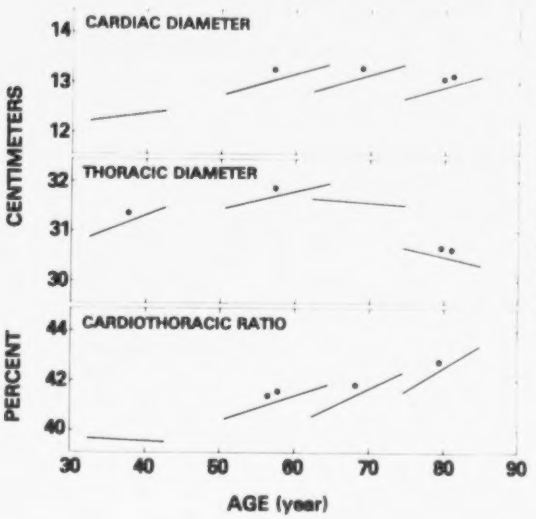


Figure 1. Longitudinal analysis of changes in radiographic variables in subjects with no evidence of heart disease. See text for explanation of construction of the line segments. Significances of slopes were compared with deviation from zero (*$P < 0.005$; **$P < 0.025$)

the older age groups than in the younger control group (Table 1). The TDs showed very little difference between the young group and the 60–69-year-old group but tended to decline in the two oldest groups compared with the young controls. The result of higher CDs and lower TDs with aging was a higher CTR in each successive age decade in both HD− (Fig. 3) and HD+ (Table 1) subjects. The distribution of CTR values by cumulative percentage curves was shifted further to the right with each successive age decade (Fig. 2). Of HD− subjects, none of those under 50 years of age had a CTR exceeding 47 per cent, but only one of the older subjects (aged 95) had a CTR exceeding 50 per cent.

The effect of normal aging, i.e., in the absence of detectable heart disease, on these variables is shown in the longitudinal analysis of chest radiographic measurements. Changes in CD, TD, and CTR (Fig. 1) generally followed the pattern suggested by the cross-sectional results, i.e., increases in CD and CTR after 50 years of age and decreases in TD in the oldest groups. The line seg-

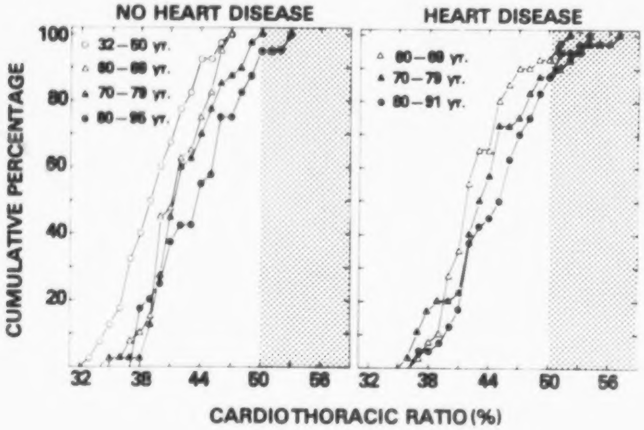


Figure 2. Distribution of cardiothoracic ratios in four age groups in subjects with and subjects without heart disease

TABLE 1
Cross-sectional Age Differences in Cardiac Diameter, Thoracic Diameter, and Cardiothoracic Ratio

	Age Range (Years)	Mean Age at Last Visit (Years)	Cardiac Diameter (cm) Mean ± SE	Thoracic Diameter (cm) Mean ± SE	Cardiothoracic Ratio (%) Mean ± SE
Men without heart disease (n = 154)	32–50 (n = 60)	43	12.5 ± 0.2	31.6 ± 0.2	39.5 ± 0.4
	60–69 (n = 32)	65	13.3 ± 0.2†	31.8 ± 0.3	41.9 ± 0.5†
	70–79 (n = 33)	75	13.4 ± 0.2†	31.4 ± 0.3	42.7 ± 0.6‡
	80–95 (n = 29)	85	13.1 ± 0.3*	30.2 ± 0.3‡	43.4 ± 0.9‡
Men with heart disease (n = 89)	60–69 (n = 29)	65	13.5 ± 0.2‡	31.4 ± 0.3	43.1 ± 0.7‡
	70–79 (n = 35)	75	13.9 ± 0.2‡	31.3 ± 0.2	44.4 ± 0.7‡
	80–91 (n = 25)	85	13.9 ± 0.3‡	30.9 ± 0.4	45.1 ± 0.8‡

Statistical comparisons were made between the youngest group with no heart disease and the older groups with and without heart disease: *$P < 0.05$; †$P < 0.005$; ‡$P < 0.001$.

ments in Figure 1 represent the rates of change in individual age groups. The slopes of CD were positive ($P < 0.025$) for the three older age groups. The slopes of TD were positive for controls and for the first older group ($P < 0.025$), but were not significant in the next age group and negative in the oldest group ($P < 0.025$). As a result, the slopes of CTR were significantly positive in all three of the older age groups. There seemed to be a discrepancy between cross-sectional and longitudinal results, in that the CD was lower in the 80–95-year-old HD– group than in the 70–79-year-old group cross-sectionally (not statistically significant) although there was an increase longitudinally.

Cross-sectional mean values for systolic blood pressure (data not shown) tended to be higher in the older age groups. However, correlation by age decade to eliminate the confounding variable of age failed to show a significant relationship between systolic blood pressure and either CD or CTR. The anthropometrically measured chest circumference (Table 2) showed trends with age similar to those for the TD, i.e., a statistically significant decline in the oldest subjects. The change in TD between the first and last visits (delta TD) was correlated ($P < 0.001$) with the change (delta) in chest circumference between corresponding visits.

Mortality Analysis

Of the original 243 subjects, 49 had died by the time this analysis was performed. Each of the 49

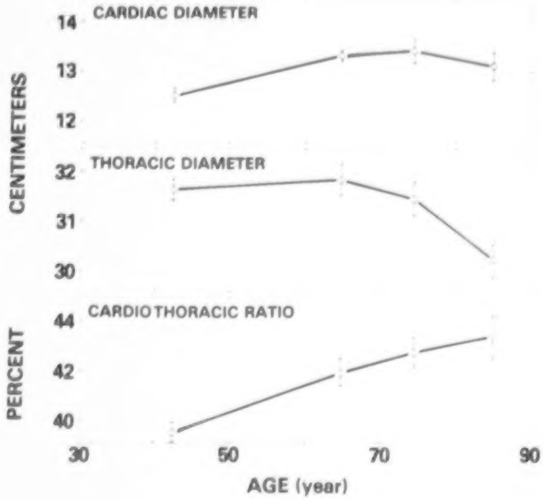


Figure 3. Cross-sectional differences in cardiac diameter, thoracic diameter, and cardiothoracic ratio in four age groups without identifiable heart disease. The cardiothoracic ratio increased with age as a result of increasing cardiac diameter and decreasing thoracic diameter.

TABLE 2

Relationship beween Anthropometrically-measured Chest Circumference and Radiographically Measured Thoracic Diameter

Mean Age at Last Visit (Years)	Chest Circumference at Last Visit (cm)	Thoracic Diameter at Last Visit (cm)	ΔChest Circumference	ΔThoracic Diameter
43 (n = 53)	101.0 ± 1.0	31.6 ± 0.2	2.2	0.6
65 (n = 49)	99.0 ± 1.3	31.6 ± 0.2	0.3	0.4
74 (n = 52)	97.0 ± 1.3†	31.3 ± 0.2	−0.7	−0.2
85 (n = 40)	93.0 ± 1.1*	30.3 ± 0.2*	−1.1	−0.4

Cross-sectional comparisons were made between the young controls and the older subjects. *$P < 0.001$; †$P < 0.05$. The correlation between changes (Δ between first and last visit) in chest circumference and thoracic diameter was significant, $r = 0.43$, $P < 0.001$.

deceased subjects was matched with a survivor in the following categories: age, time of last visit, and cardiovascular status; an additional item was the time between the last visit and death or known survival. A comparison of the mean slopes of CD and CTR for the deceased subjects and the survivors (Table 3) showed that the slopes for subjects dying with known heart disease were significantly higher than those for the survivors. However, for subjects who died without evidence of heart disease, the slopes were no different from those for matched survivors.

DISCUSSION

The possible effect of normal aging changes on the setting of normative standards for laboratory procedures has at times been neglected. Standards are sometimes set by an analysis of data from young and middle-aged adults. The determination of normality is complicated in older subjects by the high prevalence of disease states that could influence test results.

Because age-related changes are known to occur in both cardiovascular and thoracopulmonary systems, we examined the usual normative standard for the CTR and the changes in both CD and TD with aging. Previous cross-sectional studies[2–4] showed values for the CTR exceeding 50 per cent in normal old age. In those studies, higher CTRs in old age were explained by larger CDs in men and smaller TDs in women of advanced age compared with young controls.

The present study demonstrated that the usual standard of normality for the CTR is indeed appropriate for older men, when those with heart disease are carefully excluded. Our longitudinal study of women was initiated too recently to allow us to carry out a similar analysis. However, the usual normative standard for the CTR may not apply to older women, since Cowan has shown[2,3] in cross-sectional studies that the CTR increases more rapidly and reaches higher values for women than for men, because of a greater reduction in TD.

A discrepancy between cross-sectional and longitudinal results occurred in the oldest HD− group, in whom there was a cross-sectional decline in CD and a positive slope for CD longitudinally (Figs. 1 and 3). If the CD were decreasing with age, a negative slope would be expected in the longitudinal line segments. Possible explanations for this discrepancy include: 1) selective mortality in subjects with the largest hearts or, less likely, 2) a cohort difference in which the oldest cohort had always had smaller hearts than the younger cohort.

The stiffening of arterial walls with aging leads to greater vascular resistance, higher systolic blood pressure, and an increase in left ventricular work.[11] Autopsy[5] and echocardiographic studies[12] demonstrate an increase in left ventricular mass in normal old age, which is attributed to increased vascular resistance.[5] Since CD is a rough measure of left ventricular mass, and systolic blood pressure reflects part of the change in vascular resistance, we explored the relationship between these variables, but found that they were not correlated. The failure to find a relationship may result from the fact that CD and systolic blood pressure are only rough approximations of left ventricular mass and vascular resistance. A similar lack of correlation has been observed between left ventricular wall thickness on the echocardiogram and systolic blood pressure.[12]

The reduction of TD with age is apparently caused by a change in the mechanical properties of the chest wall. In a previous study,[8] it was shown that expansion of the rib cage, measured as change in chest circumference during inspiration, accounted for 39 per cent of the vital capacity in young subjects but only 30 per cent in older subjects, while abdominal wall and diaphragmatic movement accounted for the remainder. Our study showed significantly smaller chest circumferences in subjects of the 70–79-year and 80–95-year age groups than in the young controls, and thus confirmed the changes in TD measured on chest radiographs (Table 2). As expected, the changes in chest circumference and TD between the first and last visits were highly correlated ($r = 0.43$, $P < 0.001$). It

TABLE 3
Mortality Analysis

	No. of Pairs	Change in Cardiac Diameter* (mm/10 Years) Deceased Group	Survivor Group	$\Delta \pm SE$	Significance	Change in Cardiothoracic Ratio* (%/10 Years) Deceased Group	Survivor Group	$\Delta \pm SE$	Significance
Men without heart disease	22	5.4	4.2	1.2 ± 0.8	NS	2.0	2.2	0.2 ± 1.1	NS
Men with heart disease	27	6.7	0.3	6.4 ± 1.2	$P < 0.001$	1.8	0.1	1.9 ± 0.8	$P < 0.01$
TOTAL	49								

Comparisons were made between 22 deceased subjects without apparent heart disease and 22 matched survivors. Similarly, 27 deceased subjects with heart disease were compared with 27 matched survivors.

* The changes are the means of the slopes computed for each of the four subject groups.

would appear that the reduction of TD observed on chest radiographs is simply another measure of the decline in rib-cage mobility with age.

A major advantage of a longitudinal study is the ability to relate a measured variable prospectively to outcome (death or disease). Cross-sectional studies suggest that left ventricular mass increases in normal old age.[12] However, in such studies, a detrimental effect of the increasing ventricular mass cannot be excluded.

Although the CTR did not exceed the 50 per cent cut-off point in the subjects without heart disease, it is possible that subjects at the upper end of the distribution might still have a less favorable mortality rate than those with smaller CTRs. In exploring the relationship between increasing CD, CTR, and death, we found that increasing CD and CTR were predictive of death in persons with heart disease but not in persons free of heart disease (Table 3). These results suggest that an increase in left ventricular mass occurs normally with aging, is not indicative of underlying disease, and is not detrimental to survival.

This prospective study has shown that the usual normative standard for CTR (50 per cent) appears to apply to older men. The CTR increases throughout the lifespan as the result of an increase in CD after 50 years of age and a reduction of TD in advanced old age. In persons without identifiable heart disease, an increase of the CD does not affect survival adversely. The increase of CD with aging was not explained by an increase in systolic blood pressure. The decrease of TD appears to be caused by a decrease in rib-cage mobility with aging.

ACKNOWLEDGMENTS

The authors express their appreciation to Ms. Elizabeth Bannon, Mr. Howard Baldwin, and Ms. Janet Mierczak for excellent technical assistance, and to Ms. Dolores Wehner, who assisted in preparation of the manuscript.

REFERENCES

1. Danzer CS. The cardiothoracic ratio: an index of cardiac enlargement. Am J Med Sci 157:513, 1919
2. Cowan NR. The heart lung coefficient in older people. Br Heart J 21:238, 1958
3. Cowan NR. The heart lung coefficient and the transverse diameter of the heart. Br Heart J 26:116, 1964
4. Edge JR, Millard FJC. The radiographic appearances of the chest in persons of advanced age. Br J Radiol 37:769, 1964
5. Weisfeldt ML. Left ventricular function. *in*: Weisfeldt ML (ed.) The Aging Heart. New York, Raven Press, 1980, p 310
6. Mittman C, Edelman NH, Norris AH, et al. Relationship between chest wall and pulmonary compliance and age. J Appl Physiol 20:1211, 1965
7. Turner J, Mead J, Wahl ME. Elasticity of human lungs in relation to age. J Appl Physiol 25:664, 1968
8. Rizzato G, Marrazini L. Thoracoabdominal mechanics in elderly men. J Appl Physiol 28:457, 1970
9. Stone JL, Norris AH. Activities and attitudes of participants in the Baltimore Longitudinal Study. J Gerontol 21:575, 1966
10. Armitage P. Statistical Methods in Medical Research. New York, John Wiley and Sons, 1971, p 281
11. Kohn RR. Heart and cardiovascular system. *in*: Finch CE, Hayflick (eds.) The Biology of Aging. New York, Van Nostrand Reinhold, 1977, p 286
12. Gerstenblith F, Frederiksen J, Yin FCP, et al. Echocardiographic assessment of a normal adult aging population. Circulation 56:273, 1977

Journal of Gerontology
1983, Vol. 38, No. 1, 51-54

Longitudinal Age Changes in Vigilance Over an Eighteen Year Interval

Reginald E. Quilter, BSEE, Leonard M. Giambra, PhD,
and Pamela E. Benson, BS[1]

An 18-year longitudinal repeat of the Mackworth Clock vigilance experiment was conducted. Skin potential response latencies (SPRL) and reaction times were taken from 33 men during the vigilance task. The longitudinal change in this study reproduced the earlier cross-sectional relationship. The 51 to 69 year olds showed faster reaction times, the 70 to 88 year olds showed slower reaction times and, the 70 to 88 year olds detected significantly fewer targets than when 18 years younger. It was found that the greater the reduction in the percentage of targets detected, the greater the increase in reaction time. The longitudinal change noted in this study reflected previously determined cross-sectional SPRL effect. However, the SPRL outcome was equivocated by a time-of-measurement effect. It was concluded that at about age 70 years a noticeable reduction in vigilance performance occurs, and this is accompanied by a reduction in autonomic and central nervous system reactivity.

Key Words: Mackworth Clock, Reactivity, Skin potential, Response latency, Reaction time

VIGILANCE refers to a central process or state reflecting the individual's readiness to respond to specific, infrequent, and unpredictable events. Readiness to respond may be indicated by latency in voluntarily responding to the infrequent and unpredictable event (i.e., the signal or target), latency or involuntary physiological responses such as the skin potential response (SPRL), and by the proportion of signals or targets detected.

Surwillo and Quilter (1964, 1965) demonstrated a slower SPRL in older individuals together with a reduced proportion of signals or targets detected. They also found that middle-aged individuals had the fastest voluntary reaction time to signals whereas younger and older individuals had slower reaction times. The slower SPRL has been taken as an indication of the underarousal of the autonomic nervous system. The reduced proportion of targets detected may be viewed as underarousal or reduced reactivity of the central nervous system. Indeed, Mackworth (1968) has theorized that the decrement in performance in the vigilance task used by Surwillo and Quilter is the result of a decrease or disappearance of an arousal response (i.e., alpha block), in the central nervous system. The reduced vigilance performance of older individuals reported by Surwillo and Quilter (1964, 1965) may be viewed as an indication of both autonomic and central nervous system underarousal. Surwillo and Quilter used a cross-sectional design to demonstrate a difference between individuals of different ages. This study investigated 18 year age changes in vigilance performance in the participants of the Surwillo and Quilter study.

METHOD

Participants. — The sample consisted of 33 men from the Baltimore Longitudinal Study of Aging. At the second testing the age distribution was: 40 to 49 ($n = 1$), 50 to 59 ($n = 7$), 60 to 69 ($n = 10$), 70 to 79 ($n = 9$), 80 to 89 ($n = 5$), 90 to 91 ($n = 1$). All 33 men were included in correlational analyses. Only 31 men were used in variance analyses; Table 1 gives the age groupings used in these analyses. Men with visual disturbances or impairment or disease associated with muscles used in the reaction task were not tested.

Design. — Participants were first tested from 1962 to 1964; second testing occurred from 1980 to 1981. The longitudinal interval ranged from 16.9 to 18.8 years ($M = 17.8$).

One question to be answered was whether each group experienced a longitudinal change from the 1962 to 1964 session to 1980 to 1981 session. The answer was obtained, for Group 1 and Group 2 separately, by determining if the 1962 to 1964 mean was significantly different from the 1980 to 1981 mean. A second question was whether the longitudinal changes for Groups 1 and 2 were different. The rationale for this comparison was that the two age groups were expected to show different or inverse longitudinal effects. The analysis was done with a one way ANOVA between Groups 1 and 2 on longitudinal change scores. A third question was whether the cross-sectional age changes of this subset of 1962 to 1964 participants were like those observed by Surwillo and Quilter (1964, 1965); for both the 1962 to 1964 data and the 1980 to 1981 data. The analysis was done by comparing Group 1 with Group 2 holding the date of testing constant. A fourth question was whether there was anything different about the 1962 to 1964 and 1980 to 1981 testing sessions. The 1980 to 1981 testing session differed from the 1962 to 1964 testing session in three ways: (a) it was 18 years later, (b) the task was performed a second time, and (c) all participants were 18 years older. Whereas age was an undesirable confound, it was decided to hold age to a constant interval, 51 to 69 years. The analysis consisted of comparing, for 51 to 69 year olds only, the mean score in 1962 to 1964 with the mean score in 1980 to 1981. A significant difference indicated a data of testing effect. The resultant implication was that the two testing sessions were not comparable directly and any longitudinal effects might be spurious.

---

[1] Gerontology Research Center, National Institute on Aging, Baltimore City Hospitals, Baltimore, MD 21212

Apparatus. — Mackworth's Clock-Test was used. The clock was a metal box with a plain white face and a single black pointer that moved in 100 discrete steps around the face. Each step occurred after a 1-sec interval. During the 1-hour test period the pointer made 23 aberrant movements (i.e., double jumps) at long and irregular intervals. The double jumps were the targets. See Surwillo and Quilter (1964, 1965) for a complete description.

Procedure. — Each participant was given (a) instruction and demonstration, (b) practice, and (c) testing. It was the task of each participant to respond as quickly as possible when a target occurred. During the entire period of more than 1 hour the participant was seated comfortably in a bucket-type seat with his legs in a horizontal position, surrounded by a white curtain, accompanied only by a faint, droning sound. The setting and procedure duplicated closely the original experimental situation (i.e., no changes were permitted that might affect performance). For example, participants performed the task at the same time of the day as in 1962 to 1964. For a detailed procedural description see Surwillo and Quilter (1964, 1965).

Electrophysiological recording. — Beckman Ag-AgCl electrodes were used. The design of the original electrode cups was modified to accept the Beckman electrodes so that the same skin surface area was measured. Electrode placement, SPRL recording procedures, and latency measurements followed that described in Surwillo and Quilter (1964, 1965).

RESULTS

The participants were divided into two cohort groups determined by age at second testing. The age interval was set at 18 years, so that the longitudinal interval and the cross-sectional interval were nearly identical, to make for ready comparison of longitudinal and cross-sectional effects. The age intervals, sample size, means and standard deviations for the dependent measures for cohort Groups 1 and 2 at both dates of testing are given in Table 1. Analyses of variance (ANOVA) were carried out, using transformed scores, for cross-sectional, longitudinal, and time-of-testing effects for all three dependent measures. The transformations used were those suggested by Winer (1971); for propor-

Table 1. Means of Each Dependent Measure for Each Cohort at Each Date of Testing

Cohort	*n*	Age	Dates of testing	Detected targets (%)	Reaction time of detected targets (millisec)	Skin potential response latency (sec)
Group 1	17	33-51	1962-64	70.59 (19.35)	520 (157)	1.76 (.26)
		51-69	1980-81	71.10 (18.95)	493 (163)	1.98 (.18)
Group 2	14	51-69	1962-64	77.02 (10.98)	466 (113)	1.85 (.29)
		70-88	1980-81	57.76 (20.21)	574 (195)	2.17 (.45)

Note. Standard deviations in parentheses.

tions or percentages the arcsin transformation was used and for time measures the log transformation was used.

Vigilance. — The number of targets detected by each man was determined and converted to a percentage. A significant difference in the longitudinal change occurred between the two cohort groups with Group 2 detecting 19% fewer targets in 1980 to 1981 than in 1962 to 1964; Group 1 detected the same percentage of targets in 1980 to 1981 and 1962 to 1964, $F\ (1,29) = 4.30$, $p < .05$, $\omega^2 = .0964$. For Group 2 the 19% fewer targets detected was significantly different from a zero change hypothesis indicating an 18-year longitudinal reduction for 70 to 88 year olds, $F\ (1,13) = 6.86$, $p < .05$, $\omega^2 = .1730$.

Reaction time. — Surwillo and Quilter (1965) found that the middle-aged group was faster than either the young or the old groups which were equally slow; the difference was statistically significant. In the current study the 70 to 88 year olds showed a significantly slower reaction time than when 18 years younger, $F\ (1,13) = 8.40$, $p < .05$, $\omega^2 = .2090$. The longitudinal change was not significantly different for the two cohort groups, $.05 < p < .10$. The cross-sectional differences were not significant but showed the relationship found by Surwillo and Quilter (1965). The correlation between number of detected targets and reaction time was not significant at either time of testing and also when considered for each cohort group separately. A significant correlation, $r = -.40$, $p < .05$, occurred between the longitudinal change in the number of targets detected and the longitudinal change in the mean reaction time for those targets — the greater the *reduction* in targets detected, the greater the increase in reaction time. This inverse relationship continued when each cohort group was considered separately.

SPRL. — There were nonsignificant cross-sectional age effects at each time of measurement. Surwillo and Quilter (1965) found a significant cross-sectional age effect where the old group had a longer latency than the young and middle-aged groups which had identical latencies. The longitudinal changes for Group 1, $F\ (1,16) = 23.18$, $\omega^2 = .3948$, and Group 2, $F\ (1,13) = 11.12$, $\omega^2 = .2654$, were significantly, $p < .01$, different from zero. The difference in the longitudinal change between Groups 1 and 2 was not statistically significant. Also, for 51 to 69 year olds the time of testing effect was significant $F\ (1,30) = 5.15$, $p < .05$, $\omega^2 = .1152$; the latencies for Group 1 in 1980 to 1981 were longer than for Group 2 in 1962 to 1964. An additional group of 17 men 51 to 69 years of age performed the Mackworth Clock task. These men had not previously performed this task. The mean latency for this nonrepeat 1980 to 1981 cohort was 1.83, and its standard deviation was .19. When the mean of this nonrepeat cohort was compared with mean of the repeat 1980 to 1981 cohort, the nonrepeat mean was significantly smaller; for transformed latencies, $F\ (1,32) = 4.35$, $p < .05$, $\omega^2 = .0896$. When the mean of the 51 to 69 year olds tested in 1962 to 1964 was compared with this nonrepeat group it was not significantly different, $F\ (1,29) = .51$, $p < .05$. This indicates that the time-of-measurement effect was confined to the longitudinally retested men.

DISCUSSION

As measured by the percentage of targets detected, this study found an age-linked decrement in vigilance performance both within and between individuals. Furthermore, this performance decrement seems to manifest itself beginning at age 70 years. The finding of Surwillo and Quilter (1964, 1965) that the middle-aged group had a significantly faster voluntary reaction time to the targets was replicated longitudinally in this study.

For SPRL the middle-age to old-age increment was .14 sec longer than the young to middle-age increment but, the difference was not significant. However, both age groups had longer latencies 18 years later. It was demonstrated that the longer SPR latencies of the 1980 to 1981 testing for Groups 1 and 2 were confined to participants previously tested in 1962 to 1964 and was not simply the result of testing in 1980 to 1981. This pattern of outcomes might be accounted for because (a) Groups 1 and 2 actually experienced 18-year decrements in arousal or (b) previous experience on the Mackworth Clock task, even if it occurred 18 years ago, produces a relaxation of vigilance on subsequent exposures to the task. The consequences of relaxation in terms of detected targets seem to have been absent for Group 1 but not Group 2. A study is presently being carried out that is examining the effect of Mackworth Clock experience on vigilance (SPRL) level.

In summary, as measured by success in detecting targets and by increased reaction time in responding to detected targets, this study demonstrated that previously obtained decrements in vigilance in those 70 years and older can be replicated within individuals who mature to that decade.

REFERENCES

Mackworth, J. T. Vigilance, arousal, and habituation. *Psychological Review*, 1968, *75*, 308–322.

Surwillo, W. W., & Quilter, R. E. Vigilance, age, and response-time. *The American Journal of Psychology*, 1964, *77*, 614–620.

Surwillo, W. W., & Quilter, R. E. The influence of age on latency time of involuntary (Galvanic skin reflex) and voluntary responses. *Journal of Gerontology*, 1965, *20*, 173–176.

Winer, B. J. *Statistical principles in experimental design (2nd ed.)*. McGraw-Hill Book Company, New York, 1971.

Hoffmeister, F., and Muller, C. (Eds.),
Brain Function in Old Age: Evaluation of Changes and Disorders, pp. 151-159.
New York: Springer-Verlag, 1979

Temperamental Predictors of Longitudinal Change in Performance on the Benton Revised Visual Retention Test Among Seventy-Year-Old Men: An Exploratory Study

Elizabeth A. Robertson-Tchabo,[1]
David Arenberg,[2] and Paul T. Costa, Jr.[2]

"Cognition" refers to the processes by which sensory input is transformed, reduced, elaborated, stored, and retrieved. Essential to any theory of cognition is a consideration of the ways in which an individual makes use of past experiences and past reactions to increase his mastery in achieving and utilizing knowledge (Bartlett, 1932). Like many concepts in everyday use, cognitive functioning causes no problems in common usage but ceases to be straightforward when subjected to critical inspection. The study of aging and cognitive functioning has followed, not surprisingly, a number of independent traditions in the psychology of cognition. One way to study cognitive functioning is to consider general intellectual functioning as measured by standardized psychometric tests of intelligence. Another approach, following the Ebbinghaus tradition, has been concerned almost exclusively with the acquisition (learning) and retention (memory) of verbal information as demonstrated in paired-associate or serial-learning tasks or in free recall. The tradition has been to separate the study of verbal memory from the study of nonverbal activities. Others (e.g., Solyom and Barik, 1965) have studied simple conditioning or the learning of psychomotor skills, usually without reference to verbal learning (Welford, 1958). A further complicating factor in the study of learning and memory is that the processes involved are extremely complex. New information must be registered, stored, and retrieved when necessary. Any observed decline in performance with increasing age could be due to a failure in any of these processes, and many experimental psychologists would prefer to subdivide each process even further.

Studies of aging and cognitive functioning consist almost entirely of cross-sectional research designs. In such investigations, two or more adult age groups of different individuals born at different times are compared on some performance measure. A longitudinal design measures changes in performance in the same individual with increasing age. The cross-sectional approach provides measures of age differences, whereas the longitudinal approach provides measures of changes with age. At best, cross-sectional results approximate changes with age; at worst, they can be extremely misleading. (Comparisons of independent samples from the same birth cohort measured at different times, a variant of the longitudinal approach, provide estimates of mean changes but not of individual changes.)

---

[1] Department of Human Development, University of Maryland, College Park.
[2] Gerontology Research Center, National Institute on Aging, National Institutes of Health, Baltimore, Maryland.

To gain further insight into the nature of normal age-related changes in cognitive functioning, large-scale parametric investigations are essential. Such studies would provide normative data across the life span, particularly at advanced ages, for a variety of tasks. Despite the theoretical and clinical significance of such data, there are very few longitudinal studies, not all of which investigate age changes in cognitive processes. Craik (1977), in a recent review of aging and human memory, has cited the collection of these parametric data as the foremost objective for future research.

The basic question "Are there age changes in cognitive functioning?" remains controversial with respect to both findings and interpretation. One source of the discrepancy among findings is variation in the nature of tasks. Longitudinal studies of intellectual functioning as measured by intelligence tests have emphasized little or no decline in abilities with increasing age until shortly before death (e.g., Baltes and Schaie, 1974). Longitudinal studies of cognitive processes such as learning and memory have shown consistently that there are age changes late in life (Arenberg, 1983; Arenberg and Robertson-Tchabo, 1977; Gilbert, 1973). Obviously, a "no-decline" position is more appealing.

Nevertheless, Botwinick (1977), after reviewing the available literature on psychometric intelligence tests, concluded that a decline in intellectual ability is clearly an integral part of the aging process. Evidence for a decrement in performance late in life can be found even in data that have been interpreted as showing no decline. Age gradients for five subtests of the Primary Mental Abilities were reported in two studies (Schaie and Strother, 1968a, 1968b). Over the seven-year interval in both studies, the mean performance declined in all five subtests for the cohorts that had been 60 years old and over when initially measured. Similarly, Schaie and Labouvie-Vief (1974) described analyses of repeated measures of three-point data covering a 14-year period; and Schaie, Labouvie, and Buech (1973) compared three independent samples of the same birth cohort measured at different times. Again, for the cohorts aged 60 years and over when initially measured, the mean performance declined for each of the five subtests of the Primary Mental Abilities in both types of comparisons. The average declines over seven years and even over 14 years were not large; but the consistency of the declines indicated that there are age changes in intellectual performance as measured by psychometric intelligence tests.

Cross-sectional studies (Davies, 1967; Kendall, 1962; Heron and Chown, 1967) of immediate memory for geometric designs--another measure of cognitive functioning--showed age differences in performance on the Graham-Kendall test. Arenberg (1978) found longitudinal age changes in addition to cross-sectional declines in performance on a test of immediate memory for geometric designs, the Benton Revised Visual Retention Test. The tasks in these tests are predominantly non-verbal, although some subjects may choose to recode the stimuli with verbal labels.

Arenberg found substantial mean declines over a six-year interval only for the oldest group, the men over 70 when initially tested. A similar result was found in a smaller longitudinal replication sample. Furthermore, when age changes were estimated from comparisons of independent samples from the same birth cohort measured at different times, the largest decline was found for the earliest-born cohort (the oldest group). The most recent evidence of decline was found for the men initially in their 60s, who showed a small mean change when measured after a six-year interval but a substantial mean decline after a 12-year interval. In summary, these results indicate that visual memory does decline late in life even for educated, relatively healthy men.

However, it should be pointed out that this conclusion is based on group mean performance. In every age group, including the oldest, there were some individuals whose performance did not decline. The focus of the current study is on 52 men who were in their 70s when first tested on the Benton. In this group, there was a wide range of changes in the number of errors from a maximum increase of 11 to a decrease of five. These individual differences raise the important question of whether we can identify predictors of an individual's decline in memory performance. Costa and Fozard (1978) have pointed out that the ways in which stable individual differences in adult personality produce differences in characteristic ways of thinking, learning, and remembering have received little scientific study. Further, they have suggested that a consideration of such psychological processes underlying cognitive performance may help to provide a better understanding of the possibilities for intervention in cognitive problems of the elderly. The purpose of this exploratory study is to examine personality predictors and correlates of change in visual-memory performance. Although many other variables studied in the Baltimore Longitudinal Study of Aging including health status, sensory acuity, and social factors, might be related to age changes in memory performance, this presentation will be restricted to temperamental traits as possible predictors.

METHOD

Subjects

The men, participants in the BLSA (see Stone and Norris, 1966), are all volunteers who agree to come periodically to Baltimore City Hospitals for a stay of 2 1/2 days, during which numerous physiological, biochemical, and behavioral measures are obtained. These subjects are predominantly white, well-educated, and of high socio-economic status; most live in the Baltimore-Washington area. The 52 subjects in the current analyses included every participant aged 70-79 years at the time of the initial administration of the Benton Revised Visual Retention Test for whom there were both a second performance measure and an initial and a second administration of the Guilford-Zimmerman Temperament Survey.

Procedure

Benton Revised Visual Retention Test. Initially, subjects were given Form C of the Benton Revised Visual Retention Test, Administration A (Benton, 1963). Form E was administered on a subsequent visit six or more years ($\bar{x}$ = 6.5 yr) after the initial test. Each form includes ten designs. In both forms, the first two designs consist of one major geometric figure, the other eight of two major figures and a peripheral minor figure. In accordance with the standard procedure for Administration A, each design was displayed for ten seconds and then withdrawn. A subject's task was to reproduce each design from memory, with no time limit. Errors were scored according to the test manual. Each design was scored independently by two psychologists, and the infrequent disagreements were resolved by discussion or by a third psychologist. The dependent variable was the total number of errors in the ten designs of a form.

Guilford-Zimmerman Temperament Survey (GZTS). This questionnaire provides an assessment of ten traits: General Activity, Restraint, Ascendance, Sociability, Emotional Stability, Objectivity, Friendliness, Thoughtfulness, Personal Relations, and Masculinity. Each subject was given the standard GZTS instructions individually, and he completed the questionnaire during his visit to the Gerontology Research Center in Baltimore. The average test-retest interval was 6.5 years. Each of the ten scales consists of 30 items, but only "yes" and "no" responses contributed to the scale score. Douglas and Arenberg (1978) had eliminated from their analyses any scale to which a subject responded with more than three question-marks. In order to maximize the sample size in the current study, however, their exclusion rules were modified. A score on any of the 10 GZTS variables that was invalidated by the presence of more than three question-marks was prorated if it had a minimum of 15 non-question-mark responses; the corresponding scale score for the other time of measurement was also prorated.

Analyses

Multiple regression analyses were carried out with the ten scales of the GZTS as independent variables and residualized change in Benton total errors as the dependent measure. It is important to note that the criterion or dependent measure, change in Benton total errors, has been residualized; that is, it represents that part of the second measure that is uncorrelated with or independent of the initial level of total errors (see Cronbach and Furby, 1970).

RESULTS

The simple correlations between the first and second measures of the GZTS and the criterion are presented in Tables 1 and 2. Only Masculinity, General Activity, and Restraint show statistically significant simple correlations at Time 1 (see Table 1). Table 2 indicates that General Activity and Restraint do not reach statistical significance at Time 2. It can be noted that Friendliness and Emotional Stability were each correlated significantly with the criterion at Time 2 but not at Time 1. Friendliness and Emotional Stability are two of the scales that contribute to a second-order factor, Emotional Health (Guilford et al., 1976, p. 30).

Table 1. Prediction of Residual Change in Benton Total Errors from Initial Scores on the GZTS

	Multiple R	R^2	r	Beta
Masculinity	.39	.15	-.39*	-.28
General Activity	.48	.23	-.30*	-.54
Restraint	.59	.35	-.29*	-.31
Ascendance	.62	.39	.16	.25
Objectivity	.64	.41	-.26	-.29
Personal Relations	.66	.43	-.17	.21
Emotional Stability	.66	.43	-.27	.04
Thoughtfulness	.66	.43	-.01	-.03
Sociability			.18	
Friendliness			-.25	

$*p < .05$

Table 2. Correlates of Residual Change in Benton Total Errors with Six-Year Retest Scores on the GZTS

	Multiple R	R^2	r	Beta
Emotional Stability	.31	.10	-.31*	-.19
Ascendance	.39	.15	.20	.28
General Activity	.45	.20	-.25	-.34
Restraint	.50	.25	-.21	-.24
Masculinity	.56	.31	-.28*	-.23
Personal Relations	.57	.32	-.27	-.20
Objectivity	.58	.34	-.22	.34
Friendliness	.59	.35	-.28*	-.19
Thoughtfulness	.60	.35	.11	.05
Sociability			.13	

$*p < .05$

Before we turn to the results of the multiple regression analyses, there are several criteria to be considered in determining the significance of the results. First, the overall F value for the multiple regression equation must be statistically significant. Moreover, if a variable is to be considered an important predictor of change in visual memory, then its individual F value must be statistically significant. Third, the magnitude of the associated standardized Beta coefficients also represents the relative importance of a predictor variable.

Table 1 presents the predictors of residual change in Benton total errors from initial scores on the GZTS scales. The eight-variable multiple regression equation yielded a multiple correlation coefficient of .66 and accounted for 43% of the variance in the dependent variable ($F = 4.08$; $df = 8,43$; $p < .01$). When the F value for each of the individual scales is used as the criterion, only three of the variables were statistically significant predictors: Masculinity ($F = 4.64$; $df = 1,50$; $p < .05$), General Activity ($F = 13.63$; $df = 1,50$; $p < .01$), and Restraint ($F = 5.09$; $df = 1,50$; $p < .05$). As can be seen in Table 1, when the magnitudes of the associated Beta values are taken into account, the important predictor variables are General Activity, Restraint, Objectivity, Masculinity, and Ascendance.

Table 2 presents the correlates of residual change in Benton total errors from scores at the second time of measurement on the GZTS. The nine-variable multiple regression equation yielded a multiple correlation coefficient of .60 and accounted for 35% of the variance in the dependent variable ($F = 2.65$; $df = 9,42$; $p < .05$). However, when the F value for each individual correlate was used as the criterion, the only significant variable was General Activity ($F = 5.60$; $df = 1,50$; $p < .05$). When one considers the standardized Beta coefficients, which are tabulated in Table 2, the important predictor variables are Objectivity, General Activity, Ascendance, Restraint, and Masculinity.

The results of the multiple regression analysis of the correlates of residualized change in Benton total errors and residual changes in the GZTS scale scores indicated that none of the residualized GZTS measures was correlated with residualized change in Benton total errors.

By all the statistical criteria, the simple correlation coefficient, the F value for the overall regression equation, the F value for each variable independently, and the magnitude of the standardized Beta coefficients, the variables that emerged as the important predictors and correlates were General Activity, Restraint, Masculinity, and Ascendance. Moreover, General Activity, Restraint, and Masculinity were important independent variables in the regression equations for both the initial and retest GZTS.

DISCUSSION

The answer to the question whether there are non-trivial predictors of residualized change in Benton performance is yes. The consistency with which some of the scales emerge as important predictors in this exploratory study is especially encouraging. At the same time, the small sample size and the need to control for the effects of other variables (e.g., health factors) clearly require that these findings be replicated. Furthermore, the reader should be reminded that, from the evidence at hand, no causal inferences can be drawn from the observed relationship between personality traits and visual memory performance. Moreover, it is possible that levels of these personality traits would not predict maintenance of performance on other cognitive tasks. It is also likely that the personality-trait predictors of maintenance of performance would be different for women.

General Activity, Restraint, and Masculinity, the scales that predicted change in Benton performance, are stable temperamental dispositions. It was the level on the initial measurement of the GZTS scales that predicted change in Benton performance and the level at the second time of measurement that was correlated with the change in Benton performance. Change in residualized GZTS scale scores did not predict change in cognitive performance, possibly because of the high intercorrelations between the GZTS scores at the two times of measurement (the average intercorrelation for the ten scales was .73).

To gain a better understanding of the nature of the relationship between the personality-trait predictors and performance, it is important to consider the personal characteristics reflected by high scores on the General Activity, Restraint, and Masculinity scales. Men who maintained their performance were, on the average, more active, energetic, and productive (high General Activity); more responsible, serious, and self-restrained, and less impulsive (high Restraint); and more analytic and task-oriented (high Masculinity).

The GZTS is a factorially valid instrument that samples long-term normal (rather than clinical) dimensions of personality. In addition, it has been recognized for some time that the ten scales are not ten independent traits (Guilford et al., 1976). Higher-order factors have been identified: Social activity (General Activity, Ascendance, and Sociability); Introversion-extraversion (Restraint and Thoughtfulness); and Emotional health (Objectivity, Personal Relations, Emotional Stability, and Friendliness). Masculinity stands alone and is a heterogeneous scale rather than a single dimension. In addition to items reflecting an interest in masculine activities and vocations and items tapping tough versus tender-minded attitudes, 13 of the 30 items on the Masculinity scale appear to tap a dimension of emotional instability. A more univocal measure of masculinity-femininity would help to clarify this situation and possibly allow a variable from the Emotional Health cluster to emerge as a predictor. Variables from the Emotional Health domain (Emotional Stability and Friendliness) had statistically significant simple correlations with the dependent variable, change in Benton performance.

IMPLICATIONS

The manner in which stable differences in adult personality might be related to differences in characteristic ways of processing information has received little attention. However, research in cognitive styles has historically been concerned with these issues. Kogan (1973, p. 160) defined cognitive style as "individual variations in modes of perceiving, remembering, and thinking or...distinctive ways of apprehending, storing, transforming, and utilizing information." Kogan also pointed out a difference in emphasis between abilities and cognitive style; abilities involve the level of a skill, whereas cognitive style involves the manner and form of cognitive processing. Previous cross-sectional studies in the aging literature have found significant relationships between personality dimensions and task performance (Heron and Chown, 1967; Botwinick and Storandt, 1974; Costa and Fozard, 1978; Fozard and Costa, 1983), although none of these studies provides evidence as to how personality traits might influence performance.

Although personality traits that predicted maintenance of visual memory performance were identified, there is no immediate implication of the underlying process or mechanisms. There is no direct evidence that individuals with these personality characteristics approach the task or process the information in a particular manner. In fact, an interesting empirical question is whether men with high scores on General Activity, Restraint, and Masculinity, and with a low Ascendance score, encode visual information in a way characteristically different from men with the opposite pattern. It is even possible that other personality dimensions, although not significant predictors, might show important links to cognitive styles of information processing or to encoding strategies. This, too, is an important question for further research.

A number of studies of age differences in organization and memory (e.g., Hulicka and Grossman, 1967) have found that older persons do not spontaneously use mediational devices, despite the fact that when given specific instructions to do so they improve their performance. When confronted with new cognitive tasks in the laboratory, many older individuals do not have a predisposition to structure, to organize, or actively to encode the information; this may be true of non-laboratory activities as well. However, it may be that older individuals with a specific personality pattern do have such a disposition, which benefits their cognitive performance. The typical older person, who tends not to organize and encode spontaneously, may with training improve on a specific task performance; but he is unlikely to generalize a specific skill to other cognitive tasks. There are two important challenges to intervention researchers. One challenge is to devise for the elderly practical procedures (e.g., mnemonics) that have broad applicability in non-laboratory situations. The results of the present study suggest a second challenge: to develop individually tailored interventions for individuals with cognitive problems. In conclusion, we suggest that a systematic approach to the diagnosis and treatment of memory and cognitive dysfunction in the elderly will be enhanced by consideration of the individual's personal characteristics.

REFERENCES

Arenberg, D. Differences and changes with age in the Benton Visual Retention Test. *J. Geront. 33*:534-540, 1978

Arenberg, D. Memory and learning do decline late in life. *In*: Birren, J.E., Munnichs, J.M.A., Thomae, H. and Marois, M. (*Eds.*), Aging: A Challenge to Science and Society. Vol. 3. Behavioural Sciences and Conclusions. London: Oxford University Press. 1983

Arenberg, D.; and Robertson-Tchabo, E. Learning and aging. Chapt. 18 *in*: Birren, J.E. and Schaie, K.W. (Eds.), Handbook of the Psychology of Aging. New York: Van Nostrand Reinhold. 1977

Bartlett, F.C. *Remembering*. Cambridge: Cambridge University Press. 1932

Benton, A.L. *The Revised Visual Retention Test: Clinical and Experimental Applications.* New York: Psychological Corporation. 1963

Botwinick, J. Intellectual abilities. *In*: Birren, J.E. and Schaie, K.E. (Eds.), Handbook of the Psychology *of* Aging. New York: Van Nostrand Reinhold. 1977

Botwinick, J.; and Storandt, M. *Memory, Related Functions and Age.* Springfield, Ill.: Charles C. Thomas. 1974

Costa, P.T., Jr.; and Fozard, J.L. Remembering the person: Relations of individual difference variables to memory. *Exp. Aging Res., Symposium Reprint*, *4*:291-304, 1978

Craik, F.I.M. Age differences in human memory. *In*: Birren, J.E. and Schaie, K.E. (Eds.), Handbook of the Psychology *of* Aging. New York: Van Nostrand Reinhold. 1977

Cronbach, L.J.; and Furby, L. How should we measure "change"--or should we? *Psychol. Bull. 74*:68-80, 1970

Davies, A.D.M. Age and memory-for-designs test. *Brit. J. Soc. Clin. Psychol. 6*:228-233, 1967

Douglas, K.; and Arenberg, D. Age changes, cohort differences, and cultural change on the Guilford-Zimmerman Temperament Survey. *J. Geront. 33*:737-747, 1978

Fozard, J.L.; and Costa, P.T., Jr. Age differences in memory and decision making in relation to personality, abilities, and endocrine function. *In*: Birren, J.E., Munnichs, J.M.A., Thomae, H. and Marois, M. (Eds.), Agi*ng*: A Challenge to Science and Society. Vol. 3. Behavioural Sciences and Conclusions. London: Oxford University Press. 1983

Gilbert, J.G. Thirty-five-year follow up study of intellectual functioning. J. Geront. 28:68-72, 1973

Guilford, J.S.; Zimmerman, W.S.; and Guilford, J.P. The Guilford-Zimmerman Temperament Survey Handbook. San Diego: EDITS Publishers. 1976

Heron, A.; and Chown, S. Age and Function. Boston: Little, Brown. 1967

Hulicka, I.M.; and Grossman, J.L. Age-group comparisons for the use of mediators in paired-associate learning. J. Geront. 22:46-51, 1967

Kendall, B.S. Memory for designs performance in the seventh and eighth decades of life. Percept. Mot. Skills 14:399-405, 1962

Kogan, N. Creativity and cognitive style: A life span perspective. In: Baltes, P.B. and Schaie, K. W. (Eds.), Life-span Developmental Psychology: Personality and Socialization. New York: Academic Press. 1973

Schaie, K.W.; and Labouvie-Vief, G. Generational versus ontogenetic components of change in adult cognitive functioning: A fourteen-year cross-sequential study. Develop. Psychol. 10:305-320, 1974

Schaie, K.W.; Labouvie, G.V.; and Buech, B.U. Generational and cohort-specific differences in adult cognitive functioning: A fourteen-year study of independent samples. Develop. Psychol. 9:151-166, 1973

Schaie, K.W.; and Strother, C.R. The effect of time and cohort differences upon age changes in cognitive behavior. Multivar. Behav. Res. 3:259-294, 1968a

Schaie, K.W.; and Strother, C.R. A cross-sequential study of age changes in cognitive behavior. Psychol. Bull. 70:671-680, 1968b

Solyom, L.; and Barik, H.C. Conditioning in senescence and senility. J. Geront. 20:483-488, 1965

Stone, J.L.; and Norris, A.H. Activities and attitudes of participants in the Baltimore Longitudinal Study. J. Geront. 21:575-580, 1966

Welford, A.T. Ageing and Human Skill. London: Oxford University Press. 1958

Journal of Gerontology
1976, Vol. 31, No. 2, 155-163

The Effect of Age on Creatinine Clearance in Men: A Cross-Sectional and Longitudinal Study[1]

John W. Rowe, MD, Reubin Andres, MD, Jordan D. Tobin, MD,
Arthur H. Norris, MS, and Nathan W. Shock, PhD[2]

Standard true 24-hour creatinine clearance determinations were performed on 884 subjects of the Baltimore Longitudinal Study. On the basis of clinical data, subjects were placed in categories indicating the presence of specific diseases or medications which might alter glomerular filtration rate. Subjects not included in these categories were considered normal ($N = 548$). In the normals, cross-sectional analysis by 10-year age groups showed a progressive linear decline in clearance from 140 ml/min/ $1.73m^2$ at age 30 to 97 at age 80. Three or more serial clearances were obtained at 12- to 18-mo. intervals on 293 normal subjects. These longitudinal data showed an acceleration of the rate of decline in creatinine clearance with advancing age. The decrease in creatinine clearance with age seen in this study represents true renal aging and is not secondary to diseases which become increasingly prevalent in the elderly. A nomogram constructed from these data provides normative age-corrected standards for creatinine clearance.

GLOMERULAR filtration rate (GFR), estimated by the clearance of inulin, urea, or creatinine, has been shown in a number of cross-sectional studies to decline with age after maturity (Beck & Vignon, 1966; Davies & Shock, 1950; Galnares, 1970; Hansen, Kampmann, & Laursen, 1970; Hollenberg, Adams, Solomon, Rashid, Abrams, & Merrill, 1974; Lewis & Alving, 1938; Muether, Schuessler, & Sommer, 1967; Pelz, Gottfried, & Paz, 1965; Shock, 1945, 1946; Siersbaek-Nielsen, Hansen, Kampmann, & Kristensen, 1971; Stewart, 1959; van Pilsum & Seljeskog, 1958; Watkins & Shock, 1955; Wesson, 1969). Although most of these studies excluded subjects with overt renal disease, there may have existed other pathology which adversely affected renal function. This is especially pertinent in the elderly subjects studied, since they were often selected from nursing homes and the wards of general hospitals or chronic disease facilities. In addition, these studies suffer from the limitations, common to all cross-sectional studies, that selective mortality and cohort differences may significantly influence the data. No age-adjusted normal standards for creatinine clearance, which is in frequent clinical use as an estimate of GFR, are currently available.

The present study, employing both cross-sectional and serial prospective (longitudinal) analysis, on a large group of active, community-dwelling men, was performed in order to determine cross-sectional age differences and longitudinal age changes in creatinine clearance, independent of the effects of disease or medication and thereby to establish age-adjusted normative standards.

METHODS

Subjects. — All participants in the Baltimore Longitudinal Study of Aging studied between July 1, 1961, and June 30, 1971, were included in the study. Over 3,300 creatinine clearances were obtained on 884 community-dwelling volunteers ranging in age from 17 through 96 years. These men spent 2½ days in the Gerontology Research Center at 12- to 18-mo. intervals, undergoing a battery of clinical, physiologic, and psychologic tests. Subjects were generally of the middle and upper socioeconomic status, and involved in sedentary work. All participants in the longitudinal study were self-recruited. Social and demographic characteristics of the study population have been described in greater detail elsewhere (Stone & Norris, 1966).

---

[1] Clinical Physiology Branch, Gerontology Research Center, National Institute on Aging, Baltimore City Hospitals, Baltimore 21224. Reprint requests to Dr. Rowe.

[2] This work was presented at the 9th Annual Meeting of Gerontological Society, Miami Beach, Nov. 8, 1973. The authors wish to acknowledge the statistical guidance of Dr. James Schlesselman of the Biometry Branch NICHD. Elsie Beard and Marvin Yiengst provided technical assistance.

Creatinine clearance methodology. — A nonfasting serum sample for creatinine was obtained on first arrival at the Gerontology Research Center (9:00-11:00 a.m.) and a 24-hour urine collection was then begun. A fasting blood sample for serum creatinine was obtained at 8:00 a.m. the next morning, and the mean of these two determinations was used in the calculation of creatinine clearance. Creatinine in serum and urine was measured as true creatinine, using a modification of the technique of Hare (Hare, 1950). Acid tungstate filtrates of serum were treated with Lloyd's reagent and acid picrate buffer to remove noncreatinine chromogens. Creatinine was then eluted from the Lloyd's reagent with alkaline picrate and measured colorimetrically. Recovery of creatinine added to serum was 100 ± 1.7% (SD). There were no changes in creatinine methodology during the period of study reported here. All clearance values were expressed per 1.73 square meters of surface area (standard clearance).

Clinical classification. — In order to define a normal group for the study of the effects of age on creatinine clearance, a clinical classification scheme was applied to all participants in the longitudinal study on each visit. Subjects were placed in specific clinical categories on the basis of the following criteria:

(1) *Nephrolithiasis:* history or x-ray evidence of renal stone.

(2) *Urinary tract infection:* history of pyelonephritis, cystitis, or urinary tract infection of undetermined location.

(3) *Gout:* clinical evidence of gout. Subjects with asymptomatic hyperuricemia alone were not placed in this category.

(4) *Prostatectomy:* history of prostatectomy of any type, regardless of indications for surgery.

(5) *Congestive heart failure:* clinical evidence of congestive heart failure at the time of study.

(6) *Coronary heart disease:* medical history and a standard 12-lead electrocardiogram were obtained on all subjects at each visit. In addition, an exercise test (treadmill or double Master's two-step) was performed when not clinically contraindicated.

Subjects with a total of two or more points on the following criteria were placed in the coronary heart disease category:

A. Angina pectoris
    definite history = 2 points
    probable history = 1 point
B. Myocardial infarction
    definite evidence = 2 points
    probable evidence = 1 point
C. Resting electrocardiogram
    definite myocardial infarction (1.1. ×)[1] = 2 points
    probable myocardial infarction (1.2. × or 1.3. ×) = 1 point
    ischemia at rest (4.1) = 1 point

D. Post-exercise electrocardiogram (Master's or treadmill): change from no S-T-J abnormality at rest to S-T-J depression of 1.0 mm or more and S-T segment horizontal or downward sloping in any of leads I, II, aVL, aVF, VI-V6 (11.1) = 1 point

(7) *Cerebrovascular disease (CVD):* At each visit, a questionnaire was completed by the examining physician regarding the presence of any signs or symptoms of CVD. In those cases in which there was evidence on history or physical examination for CVD, a second examination was performed by another physician in order to provide a definite classification for each patient. The CVD category was composed of subjects with definite evidence for any of the following:

A. Cerebral infarction, regardless of pathogenetic mechanism,
B. Transient cerebral ischemia without infarction,
C. Chronic brain syndrome,
D. Carotid arteriosclerosis, defined as the presence of a carotid bruit on any visit or absence of carotid pulse on two or more visits.

(8) *Diabetes mellitus:* Participants in the Longitudinal Study undergo a series of four tolerance tests of carbohydrate metabolism (oral glucose, cortisone glucose, intravenous glucose, and intravenous tolbutamide) on successive visits. Performance on each test is converted to a percentile rank using nomograms developed from our study population (Andres, 1971; Pozefsky, Colker, Langs, & Andres, 1965; Swerdloff, Pozefsky, Tobin, & Andres, 1967), and categorized as previously described. The diagnosis of diabetes mellitus was made from a review of each subject's test results.

(9) *Abnormal urinalysis:* Presence of any red blood cell casts or granular casts on any visit resulted in the inclusion of a subject in this category for all of his visits. Subjects were also included if mean values on urinalysis equalled or exceeded:

protein + 1
white blood cells 5 per high power field
red blood cells 5 per high power field

(10) *Miscellaneous renal disease:* Included in this category were subjects with a definite history of previous renal disease not included in another category, i.e., history of nephrotic syndrome, glomerulonephritis, renal tumor, polycystic renal disease, renal trauma, renal or ureteral surgery, documented structural abnormalities of the genitourinary tract or diseases which commonly alter renal function including multiple myeloma, systemic lupus erythematosus, retroperitoneal fibrosis, retroperitoneal lymphoma, perirenal abscess and Fabry's disease.

(11) *Medications:* A history of medication administration was obtained on each visit. Subjects were placed in this category if they were taking any of the following medications: (a) diuretics, (b) antihypertensives, (c) digitalis preparations, (d) steroids (sex or adrenal), excluding topical use, (e) vasodilators, (f) amphetamines.

(12) *Normal group:* Subjects were placed in the normal group for those visits during which they were not included

---

[1] Number in parentheses refer to ECG reading according to World Health Organization revision of the Minnesota Code for classification of the electrocardiogram for population studies (Rose & Blackburn, 1968).

in any of the disease or drug categories. Thus, an individual may have been considered normal for all, some or none of his visits. No subjects were excluded from the normal group purely on the basis of blood pressure measurements or clinical evidence of prostatism. The effects of these variables are discussed later.

Subjects with diagnoses of diabetes mellitus, coronary heart disease, abnormal urinalysis or cerebrovascular disease at any time during their participation in the study, were considered, for the purpose of data analysis, to be in those respective categories from their entrance into the study. None of the results on these patients was included in the analysis of the normal group. Subjects in other disease or medication groups were excluded from the normal group only for those visits after the disease was diagnosed or while the medication was administered. A breakdown, by age group and disease category, of the subjects excluded from the normal group is presented in Table 1.

Simultaneous inulin and creatinine clearances. — Simultaneous inulin and creatinine clearances were performed on 55 healthy male volunteers who were not participants in the Longitudinal Study. These subjects were selected after thorough clinical evaluation which excluded those men with evidence of recent or remote hypertension, renal, cardiac, or cerebrovascular disease. Creatinine and inulin determinations were performed on identical serum and urine specimens. Inulin clearances were performed using a constant infusion technique as previously described (Davies & Shock, 1950). Inulin was measured according to the technique of Harrison (1942).

Statistical methods. — Standard statistical methods such as analysis of variance, regression analysis, and Student's t test were used. In cross-sectional analysis of the normal group each subject was represented by his mean age and the mean serum creatinine or creatinine clearance of his "normal" visits. In longitudinal analysis the annual rate of change of creatinine clearance was computed as the slope of the regression line for each subject with three of more "normal" data points.

RESULTS

Simultaneous inulin and creatinine clearances. — The mean ratio of creatinine to inulin clearance for the 55 simultaneous studies was 1.29. As is shown in Fig. 1, age had no significant effect on this ratio. Regression of the individual ratios on age yielded the equation:

$$\frac{\text{creatinine clearance}}{\text{inulin clearance}} = 1.22 + 0.0012\,\text{age (yr)}$$

$(r = 0.122)$.

Creatinine clearance and serum creatinine — cross-sectional results. — There was a highly significant reduction in creatinine clearance with advancing age in the normal subjects (N = 548). The data, when expressed as age decade means (Table 2, Fig. 2), suggested that creatinine clearance remained stable until age 34 and thereafter declined, with the rate of decline increasing after age 65. These clearance results are associated with a significant but less striking age-related increase in serum creatinine in the normal group, over the entire age-range studied ($p < 0.02$) (Table 2). The age-related de-

Table 1. Number of Subjects Excluded from Normal Group by Disease Category and Age Group.[a]

	Age Interval (Years)			
	17-24	25-44	45-64	65-96
Nephrolithiasis	—	8	47	21
Urinary tract infection	—	17	32	25
Gout	—	4	10	4
Prostatectomy	—	2	11	60
Congestive heart failure	—	—	2	3
Coronary heart disease	—	7	45	44
Cerebrovascular disease	—	2	16	18
Diabetes mellitus	—	10	36	18
Abnormal urinalysis	1	10	11	16
Miscellaneous renal disease	1	15	17	18

[a] Since some subjects had more than one disease, the total number of individual exclusions (531) in this table is necessarily larger than the number of subjects excluded (336).

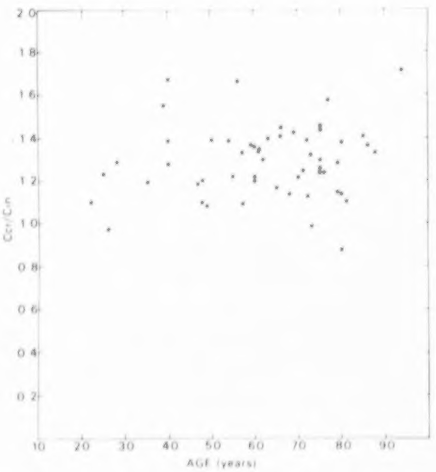


Fig. 1. Relationship between the ratio of creatinine clearance (Ccr) to inulin clearance (Cin) and age in 55 normal male subjects.

cline in creatinine excretion, which is probably secondary to declining muscle mass, accounts for the failure of serum creatinine to rise more dramatically in the face of such major changes in creatinine clearance (Table 2).

The relationship of creatinine clearance to age in 548 normal subjects (Fig. 3) was analyzed by least squares fit of linear, quadratic, and cubic polynomials. The quadratic and cubic solutions did not significantly improve the fit obtained by the linear solution, which was: Creatinine clearance (ml/min/1.73m²) = 165.57 - 0.80 age (years).

The linear equation was used to construct the nomogram shown in Fig. 4. The nomogram permits the determination of the age-adjusted percentile rank in creatinine clearance for individuals whose age and true creatinine clearance levels are known (see Appendix for details of nomogram construction).

Table 2. Cross-Sectional Age Differences in Creatinine Clearance, Serum Creatinine, and 24-Hour Creatinine Excretion.

Age (Years)	No. Subjects	Creatinine Clearance ml/min/1.73m²	Serum Creatinine Concentration mg/100 ml	Creatinine Excretion mg/24 hr
17 - 24	10	140.2[a] ±3.7	0.808 ±0.026	1790. ±52.
25 - 34	73	140.1 ±2.5	0.808 ±0.010	1862. ±31.
35 - 44	122	132.6 ±1.8	0.813 ±0.009	1746. ±24.
45 - 54	152	126.8 1.4	0.829 ±0.008	1689. ±18.
55 - 64	94	119.9 ±1.7	0.837 ±0.012	1580. ±22.
65 - 74	68	109.5 ±2.0	0.825 ±0.012	1409. ±25.
75 - 84	29	96.9 ±2.9	0.843 ±0.019	1259. ±45.

[a] Values indicate mean ± 1 SEM.

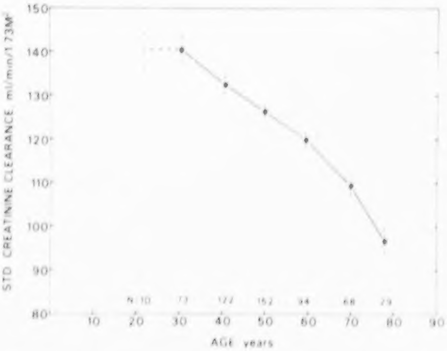


Fig. 2. Cross-sectional differences in standard creatinine clearance with age. The number of subjects in each age group is indicated above the abscissa. Values plotted indicate mean ± S.E.M.

In the longitudinal study, data were arbitrarily analyzed in 2-year study periods. The cross-sectional results by study periods are presented in Table 3. The consistency of these data, collected over a 10-year span, minimizes the possibility that methodologic changes or secular differences significantly influenced the data.

As indicated in the clinical classification scheme, no blood pressure criteria were applied in selection of the normal group. The age-adjusted partial correlation of supine basal blood pressure and creatinine clearance was: r_{partial} = 0.010 for systolic pressure vs. creatinine clearance and 0.016 for diastolic pressure vs. creatinine clearance. A partial r of 0.085 was required for significance at the 0.05 level. Therefore, there is no significant relationship between blood pressure and creatinine clearance in this carefully selected population.

Symptoms of prostatism, reflecting some degree of bladder outlet obstruction, increase in prevalence with age, and were reported in more than 80% of our subjects over 60 years old. The normal group was examined to determine if clinical findings of prostatism were related to

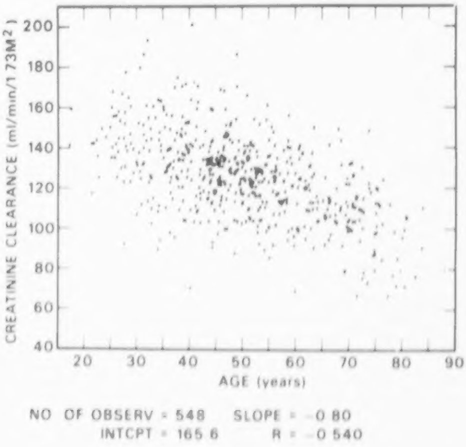


Fig. 3. Cross-sectional analysis of individual creatinine clearance vs. age. Each datum plotted represents the mean age and mean clearance values for an individual. Results of regression analysis of standard creatinine clearance on age are presented below the abscissa. Note that the intercept value of 165.6 refers to the creatinine clearance at age zero.

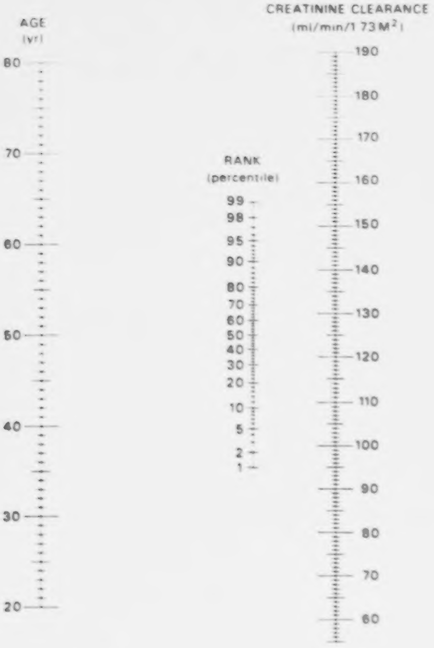


Fig. 4. Nomogram for determination of age-adjusted percentile rank in true creatinine clearance. Clearances based on total chromogen creatinine determinations using AutoAnalyzer technique may be multiplied by 1.25 to obtain equivalent true creatinine clearance for use on the nomogram (Healy, 1968).

creatinine clearance. Each subject was characterized with respect to prostate size on rectal examination (zero to 4 + enlargement) and the presence of nocturia, hesitancy, frequency, dysuria, dribbling, and reduction in force of urinary stream. Analysis of variance showed no significant effect of any of these variables, singly or in concert, on creatinine clearance.

Creatinine clearance — longitudinal results. — In longitudinal studies it is well known that the initial data point may differ significantly from subsequent results because of learning effects or because of the stressful effect of an unfamiliar testing environment. An analysis of the data in the present study failed to detect any trend for such a "first visit artefact."

Results of the longitudinal study of creatinine clearance, by age decades, are shown in Table 4. The mean clearance values approximate the cross-sectional data, suggesting that the subjects included in the longitudinal analysis were representative of the entire normal group. The slope of the regression for creatinine clearance on age was calculated for each individual with three or more "normal" data points. The similarity of the slopes, by decades, to the cross-sectional trends, is shown in Fig. 5.

Regression analysis of these individual slopes on age indicated a minimally statistically significant acceleration of the rate of decline of renal function with increasing age (Fig. 6). This trend for a greater rate of decline with age was also present but did not reach statistical

Table 3. Regression of Creatinine Clearance on Age (Data Derived from Cross-Sectional Studies Carried Out During 5 Cycles in a 10-Year Period).[a]

Study Dates[b]	1961-1963	1963-1965	1965-1967	1967-1969	1969-1971
No. subjects[c]	139	239	293	322	379
Mean creatinine clearance ml/min/1.73m²	123.5	126.2	126.7	123.5	125.5
Mean age (years)	52.4	50.2	51.6	52.4	51.5
Intercept ml/min/1.73m²	164.3	165.4	169.7	167.6	164.3
Slope ml/min/1.73m²/yr	-0.78	-0.78	-0.83	-0.84	-0.76

[a] Subjects with two visits during one cycle are represented by the age and creatinine clearance on the first visit.

[b] Study cycles ran from July 1 of the initial year to June 30 of the final year.

[c] The increase in number of subjects in succeeding study cycles reflects addition of subjects to the study.

Table 4. Longitudinal Analysis of Age-Related Changes in Creatinine Clearances.

Age (Years)	No. Subjects	Creatinine Clearance ml/min/1.73m²	Creatinine Clearance Slope ml/min/1.73m²/yr
17 - 24	1	125.3	-1.75
25 - 34	20	140.4 ±4.6	-1.09 ±0.70
35 - 44	64	132.7 ±2.0	-0.11 ±0.36
45 - 54	95	128.1 ±1.6	-0.73 ±0.30
55 - 64	60	121.8 ±1.9	-1.64 ±0.41
65 - 74	36	110.0 ±2.6	-1.30 ±0.57
75 - 84	17	97.0 ±3.4	-1.07 ±0.77
17 - 84	293	124.7 ±1.1	-0.90 ±0.18

Note: Values indicate mean ± 1 SEM

significance in the cross-sectional data (Table 2, Fig. 2).

DISCUSSION

Age is now recognized as exerting a significant influence on many physiologic characteristics. The usual design employed in gerontological studies has been cross-sectional, comparing measurements made on subjects of various ages, and attributing the differences seen to the effect of age. However, results of these studies do not necessarily reflect true age changes, since the elderly subjects represent highly selected survivors who may only have reached old age by virtue of a particular characteristic in the variable under study (selective mortality), or as the results of some environmental effect peculiar to their cohort.

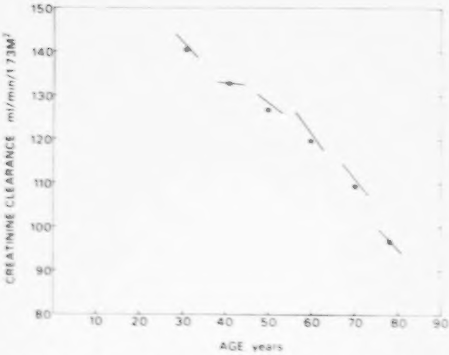


Fig. 5. Comparison of cross-sectional age differences and longitudinal age changes in creatinine clearance. The dots represent the mean values for each age decade obtained from cross-sectional data (Table 2). Longitudinal results (Table 4) are represented by line segments which indicate the mean slope of changes in creatinine clearance for each age decade. Lines are drawn with the midpoints at the mean clearance for each age decade, and with their lengths, along the abscissa, representing the mean time span over which the longitudinal data were collected for each age group.

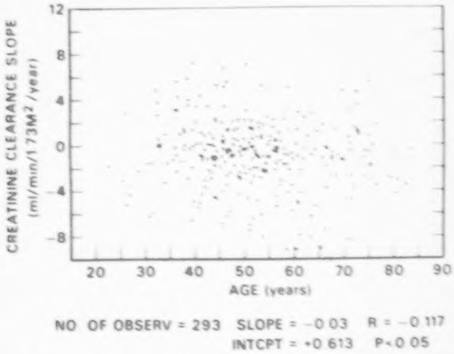


Fig. 6. Regression of individual creatinine clearance slopes on age. Computation of slopes was based on at least three data points for each subject. The slope of -0.03 ml/min/1.73 m²/yr/yr indicates that the rate of loss of creatinine clearance increases with age.

These limitations of cross-sectional studies are avoided by a longitudinal design, in which serial measurements are made in the same individuals over time. In this way, true age changes for individuals may be estimated and the rates of change for subjects of different ages compared.

In addition to study design, population selection is of prime importance in any gerontologic study. Elderly subjects are very often selected from the wards of general hospitals, chronic disease facilities, and nursing homes. Although these patients are usually screened for the presence of overt abnormalities in the organ being investigated, they too often represent a debilitated, chronically ill population. In order to avoid such a suboptimal group for testing, we have studied only active community-dwelling men and screened them with an extensive clinical evaluation, as described above.

Inulin clearance is generally recognized as the standard for measurement of GFR in man. Since the determination of inulin clearance is technically demanding, time consuming, and requires the infusion of an exogenous substance, its use has been limited to that of a research tool. Creatinine clearance, which is inexpensive, simple, and does not require administration of an exogenous substance, has been found to be clinically useful as a measure of GFR. The clearance of creatinine is independent of urine flow rate within the normal range (Chesley, 1938; Linss, Egger, & Mallman, 1970; Shaffer, 1908), and the plasma level, although influenced by diet, is relatively stable (Addis, Barrett, Poo, Ureen, & Lippman, 1951).

The factor which has most confused interpretation of creatinine clearance data has been the variety of creatinine methodologies employed. Creatinine can be measured as "true" creatinine by absorbing creatinine onto Lloyd's reagent, as in this study, or as total creatinine chromogen. The latter method measures, in addition to creatinine, noncreatinine chromogens about which little is known and which appear in the urine in very small amounts (Hare, 1950;

Haugen & Blegen, 1953; Miller & Dubos, 1937; Owen, Iggo, Scandrett, Stewart, 1954). Because of variance among individuals in the plasma levels and the low clearance of noncreatinine chromogens, total chromogen clearances will not reflect changes in GFR as clearly as, and will yield lower values than, true creatinine clearances (Healy, 1968; Relman & Levinsky, 1971).

The clearance of true creatinine has been shown in the present study, and in several others, to exceed the clearance of inulin in normal subjects, thus indicating tubular secretion of creatinine (Breckenridge & Metcalfe-Gibson, 1965; Doolan, Alpen, & Theil, 1962; Healy, 1968; Jeremy & McIver, 1966; Mandel, Jones, Willis, & Cargill, 1953; Miller & Winkler, 1938). The results of our simultaneous inulin and creatinine clearances show that the renal handling of creatinine, relative to inulin, does not change with age. Therefore the age-related decline in creatinine clearance represents a reduction in GFR and cannot be attributed to a decrease in renal tubular secretion of creatinine.

Duration of urine collection for creatinine clearance varies in published reports from 1 to 24 hours. Twenty-four hour collections seem most useful since they avoid the effect of diurnal variation in GFR (Wesson, 1969), distribute the influence of activity and diet over a complete day, and minimize the impact of the inevitable collection errors.

Our results demonstrate a highly significant age-related decline in creatinine clearance in normal men and establish age-adjusted normal standards. To facilitate the use of these standards, a nomogram has been developed (Fig. 4) which enables the determination of an individual's age-adjusted percentile rank in creatinine clearance. Clearances based on the commonly used AutoAnalyzer creatinine determinations for total chromogens may be converted to equivalent true creatinine clearances by multiplying AutoAnalyzer clearance by 1.25 (Healy, 1968).

In spite of the voluminous literature on creatinine clearance, very few studies of normal subjects have been reported which are directly comparable to the present report, that is, which collected urine over 24 hours, which measured true creatinine, and which recorded age, sex, and surface area. Those studies that are available, although on small numbers of subjects, are in general agreement with this study (Doolan, 1962; Miller & Winkler, 1938; van Pilsum & Seljeskog, 1958).

No adverse effect of elevated blood pressure on GFR was detectable in this study. Our subjects have ready access to medical care and reports of their test results are sent to their private physicians after each visit. Thus, men who are consistently hypertensive are likely to be started on therapy and thereby to be excluded from the normal group for subsequent visits. The data presented in this study therefore do not reflect the effect of sustained hypertension on GFR. A similar mechanism may be operative in our inability to detect a significant effect of prostatism on GFR. The highly educated subjects in this study are likely to seek medical attention prior to developing symptoms reflecting bladder obstruction of sufficient magnitude to result in impaired renal function.

The mean decline in creatinine clearance among the different 10-year age groups has been well defined by these studies. The decline in clearance for individual subjects however can not be defined with confidence over the time period (mean of 6 years) covered by these studies. The problem of experimental design for longitudinal studies (duration of study, frequency of testing) has been recently discussed by Schlesselman (1973). We can compute from that report that, in order to define an individual's slope with minimally acceptable accuracy, we would require annual testing for 18 years for normal subjects. Other experimental strategies (even more frequent testing for shorter periods of time) are possible; we wish to emphasize that studies of age changes in creatinine clearance should not be undertaken lightly.

The decline in GFR with age may re related to several physiologic and pathologic changes seen in the senescent kidney. A progressive reduction in renal plasma flow (RPF) with advanding age seems well established (Wesson, 1969). Factors contributing to this decreased RPF include an age-related decline in cardiac output (Brandfonbrener, Landowne, & Shock, 1955; Krovetz & Goldbloom, 1972; Lammerant, Veall, & DeVisscher, 1961) and reductions in the renal vascular bed (Ljungqvist & Lagergren, 1962; Moore, 1931; Oliver, 1952; Sworn & Fox, 1972; Takazakura, Sawabu, Handa, Takada, Shinoda, & Takeuchi, 1972).

Ljungqvist and Lagergren (1962), in a postmortem study employing both micro-angiographic and histologic techniques, detected an age-related increase in the number of blind aglomerular arterioles, particularly in the cortex, and concluded that a significant portion of blood flow to the senile kidney is shunted from cortical to medullary areas. Takazakura et al., 1972, using similar techniques, have confirmed these findings. Recently, Hollenberg et al., 1974, employing the xenon washout technique, demonstrated that the age-related decline in renal blood flow primarily reflects decreased perfusion of the renal cortex.

To our knowledge, the present study represents the first longitudinal study of GFR in man. Since the longitudinal changes approximate the cross-sectional differences, there does not appear to be any significant effect of selective mortality or of differences among cohorts in the cross-sectional results.

APPENDIX

Construction of the nomogram. Nomogram construction was based on the following observations in the normal group:

1. Endogenous creatinine clearance decreases with increasing age. $y = 165.57 - 0.80x$ where y equals creatinine clearance in ml/min/1.73m^2 surface area and x equals age in years.
2. The individual observations are normally distributed about this regression line.
3. The variance in creatine clearance does not change significantly with age (Table 2).

Percentile rank, as it relates to standard deviation in a normally distributed population, was obtained from a Table of Probit Transformation. The creatinine clearance level for selected percentile rankings at ages 30, 50, and 70 were computed. Thus the 50th percentiles (means) were computed from formula 1 above. The 1st and 99th percentiles were computed from the mean values minus and plus 2.33 standard deviations. Other rankings were similarly computed. Nomogram construction was then begun by arbitrary placement of suitable scales for age and for creatinine clearance on a sheet of paper. Fiftieth centile lines drawn from the 3 selected ages to the three appropriate creatinine clearances, intersected each other at a point which fixed the 50th centile point of the percentile rank scale (see Fig. 4). Other locations on this scale were similarly fixed.

REFERENCES

Addis, T., Barrett, E., Poo, G., Ureen, H. J., & Lippman, R. W. The relation between protein consumption and diurnal variations of the endogenous creatinine clearance in normal individuals. *Journal of Clinical Investigation*, 1951, *30*, 206-209.

Andres, R. Aging and diabetes. *Medical Clinics of North America*, 1971, *55*, 835-846.

Beck, H., & Vignon, J. C. Clearance de la créatinine endogène, chez les personnes âgées. *Revue Francaise de Gérontologie*, 1966, *12*, 145-148.

Brandfonbrener, M., Landowne, M., & Shock, N. W. Changes in cardiac output with age. *Circulation*, 1955, *12*, 557-566.

Breckenridge, A., & Metcalfe-Gibson, A. Methods of measuring glomerular filtration rate: A comparison of inulin, vitamin B-12, and creatinine clearances. *Lancet*, 1965, *2*, 265-267.

Chesley, L. C. Renal excretion at low urine volumes and the mechanism of oliguria. *Journal of Clinical Investigation*, 1938, *17*, 591-597.

Davies, D. F., & Shock, N. W. Age changes in glomerular filtration rate, effective renal plasma flow, and tubular excretory capacity in adult males. *Journal of Clinical Investigation*, 1950, *29*, 496-507.

Doolan, P. D., Alpen, E. L., & Theil, G. B. A clinical appraisal of the plasma concentration and endogenous clearance of creatinine. *American Journal of Medicine*, 1962, *32*, 65-79.

Galnares, J. S. Volúmen del filtrado glomerular en el anciano. *Revista Española de Gerontologia*, 1970, *5*, 1-6.

Hansen, T. M. Kampmann, J., & Laursen, H. Renal excretion of drugs in the elderly. *Lancet*, 1970, *1*, 1170.

Hare, R. S. Endogenous creatinine in serum and urine. *Proceedings of the Society for Experimental Biology & Medicine*, 1950, *74*, 148-151.

Harrison, H. E. Modification of the diphenylamine method for determination of inulin. *Proceedings of the Society for Experimental Biology & Medicine*, 1942, *49*, 111-114.

Haugen, H. N., & Blegen, E. M. The true endogenous creatinine clearance. *Scandinavian Journal of Clinical & Laboratory Investigation*, 1953, *5*, 67-71.

Healy, J. K. Clinical assessment of glomerular filtration rate by different forms of creatinine clearance and a modified urinary phenolsulphonphthalein excretion test. *American Journal of Medicine*, 1968, *44*, 348-358.

Hollenberg, N. K., Adams, D. F., Solomon, H. S., Rashid, A., Abrams, H. L., & Merrill, J. P. Senescence and the renal vasculature in normal man. *Circulation Research*, 1974, *34*, 309-316.

Jeremy, D., & McIver, M. Inulin, ^{57}Co-labelled vitamin B-12 and endogenous creatinine clearances in the measurement of glomerular filtration rate in man. *Australasian Annals of Medicine*, 1966, *15*, 346-351.

Krovetz, L. J., & Goldbloom, S. Normal standards for cardiovascular data. I. Examination of the validity of cardiac index. *Johns Hopkins Medical Journal*, 1972, *130*, 174-186.

Lammerant, J., Veall, N., & DeVisscher, M. Observations on cardiac output and "pulmonary blood volume" in normal man by external recording of the intracardiac flow of ^{131}I labelled albumin. *Nuclear Medicine*, 1961, *1*, 353-379.

Lewis, W. H., Jr., & Alving, A. S. Changes with age in the renal function in adult men. I. Clearance of urea. II. Amount of urea nitrogen in the blood. III. Concentrating ability of the kidneys. *American Journal of Physiology*, 1938, *123*, 500-515.

Linss, G., Egger, E., & Mallman, N. W. Einige Aspekte zur Brauchbarkeit der endogenen Kreatininclearance. *Deutsche Gesundheitswesen*, 1970, *25*, 2173-2178.

Ljungqvist, A., & Lagergren, C. Normal intra-renal arterial pattern in adult and aging human kidney. *Journal of Anatomy*, 1962, *96*, 285-300.

Mandel, E. E., Jones, F. L., Willis, M. J., & Cargill, W. H. Renal excretion of creatinine and inulin in man. *Journal of Laboratory & Clinical Medicine*, 1953, *42*, 621-637.

Miller, B. F., & Dubos, R. Determination by a specific enzymatic method of the creatinine content of blood and urine from normal and nephritic individuals. *Journal of*

Biological Chemistry, 1937, *121*, 457-464.

Miller, B. F., & Winkler, A. W. The renal excretion of endogenous creatinine in man; comparison with exogenous creatinine and inulin. *Journal of Clinical Investigation*, 1938, *17*, 31-40.

Moore, R. A. The total number of glomeruli in the normal human kidney. *Anatomical Record*, 1931, *48*, 153-168.

Muether, R. O., Schuessler, W. P., & Sommer, A. J. Laboratory studies on the aging kidney. *Journal of the American Geriatric Society*, 1967, *15*, 260-275.

Oliver, J. R. Urinary system. In A. I. Lansing (Ed.), *Cowdry's problems of aging*. Williams & Wilkins, Baltimore, 1952.

Owen, J. A., Iggo, B., Scandrett, F. J., & Stewart, C. P. The determination of creatinine in plasma or serum, and in urine: a critical examination. *Biochemical Journal*, 1954, *58*, 426-437.

Pelz, K. S., Gottfried, S. P., & Paz, E. Kidney function studies in old men and women. *Geriatrics*, 1965, *20*, 145-149.

Pozefsky, T., Colker, J., Langs, H., & Andres, R. The cortisone-glucose tolerance test. The influence of age on performance. *Annals of Internal Medicine*, 1965, *63*, 988-1000.

Relman, A. S., & Levinsky, N. E. Clinical examination of renal function. In M. B. Strauss & L. G. Welt (Eds.), *Diseases of the kidney*. Little, Brown, Boston, 1971.

Rose, G. A., & Blackburn, H. *Cardiovascular survey methods*. World Health Organization, Geneva, 1968.

Shaffer, P. The excretion of kreatinine and kreatine in health and disease. *American Journal of Physiology*, 1908, *23*, 1-10.

Shock, N. W. Inulin, diodrast and urea clearance studies on aged human subjects. *Federation Proceedings*, 1945, *4*, 65.

Shock, N. W. Renal function tests in aged males. *Geriatrics*, 1946, *1*, 232-239.

Schlesselman, J. J. Planning a longitudinal study: II. Frequency of measurement and study duration. *Journal of Chronic Diseases*, 1973, *26*, 561-570.

Siersbaek-Nielsen, K., Hansen, J. M. Kampmann, J., & Kristensen, M. Rapid evaluation of creatinine clearance. *Lancet*, 1971, *1*, 1133-1134.

Stewart, C. P. Renal function in the aged. *Gerontologia Clinica*, 1959, *1*, 160-167.

Stone, J. L., & Norris, A. H. Activities and attitudes of participants in the Baltimore Longitudinal Study. *Journal of Gerontology*, 1966, *21*, 575-580.

Swerdloff, R. S., Pozefsky, T., Tobin, J. D., & Andres, R. Influence of age on the intravenous tolbutamide response test. *Diabetes*, 1967, *16*, 161-170.

Sworn, M. J., & Fox, M. Donor kidney selection for transplantation. *British Journal of Urology*, 1972, *44*, 377-383.

Takazakura, E., Sawabu, N., Handa, A., Takada, A., Shinoda, A., & Takeuchi, J. Intrarenal vascular changes with age and disease. *Kidney International*, 1972, *2*, 224-230.

van Pilsum, J. F., & Seljeskog, E. L. Long term endogenous creatinine clearance in man. *Proceedings of the Society for Experimental Medicine & Surgery*, 1958, *97*, 270-272.

Watkins, D. M., & Shock, N. W. Agewise standard value for C_{in}, C_{pah}, Tm_{pah} in adult males. *Journal of Clinical Investigation*, 1955, *34*, 969.

Wesson, L. G., Jr. Renal hemodynamics in physiological states. In L. G. Wesson, Jr. (ed.) *Physiology of the human kidney*. Grune & Stratton, New York, 1969.

J Chron Dis 1973, Vol. 26, pp. 553–560.

PLANNING A LONGITUDINAL STUDY: I. SAMPLE SIZE DETERMINATION

JAMES J. SCHLESSELMAN

Biometry Branch,
National Institute of Child Health and Human Development,
Bethesda, Md. 20014, U.S.A.

1. INTRODUCTION

I WISH to explain how the use of some elementary statistical techniques can aid in the planning and continuing evaluation of a longitudinal study. Though the methods are well known as matters of principle, an exposition of their specific application to longitudinal studies may aid investigators by providing explicit examples of their use.

In the course of evaluating the longitudinal study of aging conducted by the Gerontology Research Center of the National Institute of Child Health and Human Development, the investigators reviewed their choices for the size of the study population, the frequency of repeated measurements on the subjects, and the length of time over which the study would continue. Originally these decisions were made in part on the basis of subject matter knowledge and expertise. The usual constraints of limitations of staff and funds, and the problems of managing such a study over an extended period of time were additional considerations which tempered the final choices. The techniques which are now being applied to the evaluation of these decisions can be used by investigators who are only in the planning stages of a longitudinal study.

An important aspect of the statistical techniques to be illustrated is that their application requires a precise statement of the goals of the study. Verbal adeptness and limpid prose will not suffice in this matter. One must be specific about one's goals in quantitative terms.

To focus discussion, I will assume that the major goals of a longitudinal investigation include:

(a) determining means and average rates of change on a number of variables for designated age groups, and detecting differences among these age groups;

(b) determining individual levels and rates of change in the study participants, and characterizing the magnitude of individual differences.

Of course one may be interested in groups other than those based upon age. For example, one may have a multiple crossed-classification based upon various clinical, sociological, and treatment factors. The use of age groups throughout this discussion is intended to permit an unencumbered exposition of principles.

2. SAMPLE SIZES FOR ESTIMATING GROUP MEANS

Two simple techniques for planning sample size to estimate group means with a pre-specified precision involve using the standard error of the mean or the coefficient of variation of the mean.

Consider a group of N subjects with measurements $x_1, x_2, \ldots, x_N$ on some variable X. The sample mean $\bar{x}$ and the sample standard deviation S are calculated as:

$$\bar{x}=\{x_1+x_2+\ldots+x_N\}/N$$

$$S=\{\sum_{i=1}^{N}(x_i-\bar{x})^2/(N-1)\}^{\frac{1}{2}}.$$

In the sample, the *standard error of the mean* $\bar{x}$ is given by $S/\sqrt{N}$. If one requires that the standard error of the mean be equal to a specified quantity, then $S/\sqrt{N}=\varepsilon$ implies that

$$N=S^2/\varepsilon^2. \tag{1}$$

To use this formula for planning purposes, one must specify ε and an *anticipated* value for S. Since the 95 per cent confidence interval based on the mean $\bar{x}$ is approximately $\bar{x}\pm 2S/\sqrt{N}$, one can regard the quantity 4ε as being approximately the width of the 95 per cent confidence interval.

The *coefficient of variation of the mean* $\bar{x}$ is given by $CV=S/(\bar{x}\sqrt{N})$. Therefore

$$N=S^2/(\bar{x}CV)^2 \tag{2}$$

gives an alternative formula for sample size. To use this formula for planning purposes, one must specify the desired coefficient of variation CV and *anticipated* values for the mean $\bar{x}$ and the standard deviation S. The coefficient of variation of the mean has a simple interpretation, namely, the sample mean $\bar{x}$ is likely ($p\approx 0.68$) to be within CV units of the true mean μ. That is, for approximately normally distributed data

$$\mathrm{Prob}\{|(\bar{x}-\mu)/\mu|<CV\}\approx 0.68.$$

Ordinarily one speaks of $CV\times 100$ per cent, percentage deviations from the true mean.

The use of formulas (1) and (2) requires specifying anticipated values for $\bar{x}$ and S. The conventional course of action in this situation is to use values based upon previously published data, a pilot study, or to use subject matter knowledge and expertise in guessing values. In the evaluation of a study in progress, the completed portion of the study serves the purpose of providing these values.

To illustrate the application of these techniques, Table 1 shows data on heat production (cal/sq m/hr) and oxygen uptake (cc/sq m/min) reported by Shock and Yiengst [1]. Corresponding to each age grouping are the reported values of $\bar{x}$ and S. Using these numbers as anticipated values in formula (2), I calculated estimates of sample size needed to attain coefficients of variation of the mean equal to 2 and 5 per cent. All calculated sample sizes have been rounded upwards. For example, for heat production in the 40–49 yr age group

$$N=(3.95)^2/(35.73\times 0.02)^2\approx 31$$

for a *CV* of 2 per cent, and

$$N=(3.95)^2/(35.73\times 0.05)^2\approx 5$$

for a *CV* of 5 per cent. Thus if one were planning a study and wished to determine the average heat production in males 40–49 yr of age with a *CV* of 2 per cent, one would choose a sample of 31 subjects. Similar interpretations apply to the other entries in Table 1. Though the estimated sample sizes agree well age by age on both variables, one should not expect this to happen as a rule. Most likely wide discrepancies in estimated sample size will occur if a number of variables are to be measured. One must choose some compromise value in this situation. Ranking in importance the variables under study along with their corresponding sample sizes may aid in the selection. One would wish to choose a value for *N* which would suffice at least for variables of high importance. Furthermore, one may wish to make adjustment for anticipated loss of participants over the course of the study.

TABLE 1. ESTIMATED SAMPLE SIZES TO ATTAIN COEFFICIENTS OF VARIATION OF THE MEAN OF 2 AND 5 PER CENT BY SELECTED AGE GROUPS, BASED UPON THE DATA OF SHOCK AND YIENGST [1]

Heat production (cal/sq m/hr)

			Sample sizes	
Age	$\bar{x}$	S	$CV=2\%$	$CV=5\%$
40–49	35.73	3.95	31	5
50–59	34.50	4.47	42	7
60–69	33.00	3.60	30	5
70–79	32.60	3.72	33	6
80–90	30.05	4.65	60	10

Oxygen uptake (cc/sq m/min)

			Sample sizes	
Age	$\bar{x}$	S	$CV=2\%$	$CV=5\%$
40–49	123.45	14.56	35	6
50–59	121.88	18.20	56	9
60–69	112.88	13.02	34	6
70–79	113.08	13.02	34	6
80–90	104.68	16.10	60	10

An important point to be remembered in any sample size calculation is that the anticipated values used in formulas (1) and (2), and in subsequent formulas, should refer to the target population. For example, means and standard deviations derived from data on restricted groups, such as patients in a hospital ward or the inhabitants of a single city, might be vastly different from the general population. If one were planning a study of the general population, such data could be quite misleading. This is one of a number of reasons why sample size calculations can be regarded as only approximate.

Another point is that the precision actually attained in a study may differ from that which was planned. This may occur because the values for $\bar{x}$ and S used in formulas (1) and (2) are only 'guesses' as to what will be obtained. This in no way invalidates the inferences to be drawn from the study, the sample estimates, or their measures of error.

The choice of whether to use formula (1) or (2) for estimating sample size is somewhat a matter of preference for ε or CV. An absolute measure of variation is given by ε, whereas CV is a relative measure. The preference within the field of survey sampling is for CV. Kish [2] gives a good discussion of the coefficient of variation and the problems of planning survey samples. He cautions that in some situations CV is best avoided. (i) If the mean of the variable is close to zero, the estimated sample sizes in formula (2) can be very erratic. (ii) For binomial variables, $\bar{x}$ corresponds to P or Q. Thus the sample size calculated from formula (2) will differ, depending on the arbitrary decision of which side of the binomial cut is regarded as P and which as Q.

From formulas (1) and (2) one sees that the relationship between ε and CV is given by $\varepsilon = \bar{x}\ CV$. Thus one can easily relate values of ε to those of CV.

3. SAMPLE SIZE FOR DETECTING DIFFERENCES IN GROUP MEANS

Having discussed two techniques for determining sample size when the goals of a study included estimating group means with pre-specified precision, I will now illustrate techniques for calculating sample size when the goals include detecting differences among the group means. To simplify the discussion I will assume that we wish to allocate equal numbers of observations to the groups, that the groups are independent, that the standard deviations within groups are approximately equal, and that the observations are approximately normally distributed. A change in any one of these assumptions will lead to a different formula for calculating sample size.

The statistical tradition of testing for differences assumes that one wishes to guard against two types of errors: Type I—claiming that a difference among groups exists, when in fact it does not; Type II—claiming that no difference exists among the groups, when in fact it does. The probability of a Type I error is denoted by α and the probability of a Type II error is designated β. The quantity α is referred to as the *level of significance* of the test. The quantity $(1-\beta)$ is called the *power* of the test. Since either type of error is undesirable, we want both α and β to be small.

To calculate a sample size to detect a difference among means, one must specify four quantities: (i) how large a difference one wishes to detect, Δ; (ii) the level of significance, α; (iii) the chance of not detecting a difference of Δ units, β; (iv) and the standard deviation, σ. In the case where one is testing for a difference between two means, a commonly used approximate formula for calculating a sample size N *for each group* is given by

$$N = 2\sigma^2(Z_\alpha + Z_\beta)^2/\Delta^2. \tag{3}$$

The quantities Z_α and Z_β are unit normal deviates corresponding to the level of significance, α and the Type II error, β. Table 2 gives values for Z_α and Z_β for a range of values of α and β. The deviates Z_α and Z_β correspond to the probability in the upper tail of the unit normal distribution.

As an example, suppose we want a 95 per cent chance of detecting a difference of 2 cal/sq m/hr in heat production between 40-yr and 50-yr old men and that we want

TABLE 2. UNIT NORMAL DEVIATES Z_α AND Z_β CORRESPONDING TO UPPER TAIL PROBABILITIES α AND β FOR ONE-SIDED AND TWO-SIDED TESTS OF SIGNIFICANCE: VALUES FOR Z_β ARE THE SAME FOR ONE-SIDED AND TWO-SIDED TESTS

$\alpha(/\beta)$	One-sided test $Z_\alpha(/Z_\beta)$	Two-sided test Z_α
0.001	3.09	3.29
0.005	2.58	2.81
0.01	2.33	2.58
0.025	1.96	2.24
0.05	1.64	1.96
0.10	1.28	1.64
0.20	0.84	1.28
0.30	0.52	1.04

differences to be significant at the 1 per cent level. Then $\beta=0.05$, $\Delta=2$ and $\alpha=0.01$. As an anticipated value for σ, we can use the pooled standard deviation $\{[26(3.95)^2+26(4.47)^2]/(26+26)\}^{\frac{1}{2}}=4.22$. (Shock and Yiengst [1] used $N=27$ in the 40- and 50-yr age groups.) Using these values along with Table 2 and formula (3), we have for a two-sided test of significance

$$N=2(4.22)^2\,(2.58+1.64)^2/(2)^2\approx 159.$$

If one wished to detect a true increase (decrease) of 2 units, then the values of Z_α and Z_β from Table 2 corresponding to a one-sided test of significance give

$$N=2(4.22)^2\,(2.33+1.64)^2/(2)^2\approx 141.$$

One should appreciate the contributions of Δ and σ in formula (3). Values specified for Δ should correspond to differences of clinical or biological *importance*. This is ultimately a subject matter judgement for which statistical theory can usually provide little guidance. By specifying small values of Δ, one can easily produce the requirement of exceedingly large samples. Conducting a study based on inflated sample sizes can result in statistically significant differences which are clinically inconsequential.

One may regard the standard deviation within groups, σ, as consisting of two components: (i) inherent biological variation among the organisms, σ_B, and (ii) measurement error, σ_E, where $\sigma^2=\sigma_B^2+\sigma_E^2$. Ordinarily one hopes that σ_E^2 is small relative to σ_B^2. Poor measurements can easily inflate σ^2, and consequently lead to requirements of larger sample sizes. One can decrease σ^2 by directly reducing σ_E^2 through improved measurement, or by replicating determinations on each subject. If one uses the mean of K replicates as the datum for each subject, σ^2 becomes $\sigma^2=\sigma_B^2+(\sigma_E^2/K)$. The relative advantages of directly reducing σ_E^2, increasing the number of replicates K, or increasing the sample size N are important issues, but will not be treated here. Refer to Cox [3] for a complete discussion.

Formula (3) shows that for a given Δ and σ, the three quantities N, α, and β are interrelated. For fixed N, one can decrease α only by increasing β, and vice versa. For example, suppose we wish to detect a difference of $\Delta=2$ between 40- and 50-yr old men, but that we can afford samples no larger than $N=50$ for each group. Using formula (3) with $\sigma=4.22$, we have

$$Z_\alpha+Z_\beta=\{50(2)^2/[2(4.22)^2]\}^{\frac{1}{2}}=2.37.$$

If we wish to make a two-tailed test at the $\alpha=0.01$ level of significance, then $Z_\alpha=2.58$, which makes $Z_\beta=-0.21$. From tables of the normal distribution function (e.g. Pearson and Hartley [4]) one finds that $\beta=0.58$. If one wishes $\alpha=0.05$, then $Z_\alpha=1.96$, giving $Z_\beta=0.41$ and $\beta=0.34$. If one wishes $\alpha=0.20$, then $Z_\alpha=1.28$, giving $Z_\beta=1.09$ and $\beta=0.14$. Thus for a fixed sample size N and fixed values of Δ and σ, one can decrease β only by increasing α. This example illustrates the trade-off that one can make between Types I and II errors.

If the size of the sample is completely at our disposal, we can reduce α and β to any preassigned level by using a large enough N. If the sample size is fixed or limited by practical and economic constraints, and if committing a Type I error is more serious than committing a Type II error, there is a case for reducing α at the expense of increasing β. In many kinds of studies one wishes primarily to characterize phenomena and describe the physical and biological world. In such instances I would be inclined to treat both types of errors on an equal footing, choosing $\alpha=\beta$.

There are other important interrelationships among the quantities N, Δ, σ, α and β, the most important being the *operating characteristic* or *power-curve*. This is simply a graph of β or $(1-\beta)$ against Δ for fixed N, σ and α. One may consult Dixon and Massey [5] as one of many elementary discussions. The usual statistical discussion speaks of testing hypotheses, a null hypothesis, and an alternative hypothesis. Within this context I have assumed the null hypothesis to be 'no difference in group means', and the alternative hypothesis to be 'a difference of Δ units in the group means'.

If one wishes to detect a difference in proportions between two groups, the commonly used approximate formula for determining the sample size N *for each group* is

$$N=\{Z_\alpha\sqrt{2p_1(1-p_1)}+Z_\beta\sqrt{p_1(1-p_1)+p_2(1-p_2)}\}^2/\Delta^2 \tag{4}$$

where p_1 is the anticipated proportion in group 1 (assumed to be the 'reference' or 'control' group) and $p_2=p_1+\Delta$ is the anticipated proportion in group 2.

Formulas (3) and (4) apply to situations in which one wishes to compare two groups. If one contemplates comparisons among several groups, these formulas could be applied for each comparison. The issue of making repeated tests of significance leads to the so-called multiple comparisons problem. A simple case of this is given by the investigator who makes 100 tests of significance at the $\alpha=0.05$ level. If the tests are independent and if there are in fact no differences between the groups compared, one would expect 5 'statistically significant' results to occur due to chance alone. There is, in fact, approximately a 95 per cent chance that between 1 and 9 tests would result in being called statistically significant at the 0.05 level. One may refer to O'Neill and Wetherill [6] for a recent survey of issues and practice with regard to multiple comparisons.

When one has several or more groups, an alternative to making repeated pair-wise tests of significance is to use the analysis of variance F-test, which is designed to detect any difference, wherever it may occur, among the groups. One disadvantage is that if a significant F-value obtains, all that one knows is that a difference exists, but the test does not say where. If one wants to know which groups are contributing to the significant F, he must retreat to pair-wise tests, partitioning degrees of freedom via orthogonal contrasts, or multiple comparisons procedures. In spite of this, the reader may find the tables provided by Kastenbaum *et al* [7–9] to be extremely useful for planning sample size.

To this point discussion has focused on estimating group means and detecting differences among them. Thus aspects of cross-sectional as opposed to longitudinal analysis have been considered. If one were contemplating a cohort analysis, say comparing a group of persons at two different points in time, formulas (3) and (4) would not be appropriate for determining sample size. The basic assumption of independent groups would be violated. The corresponding approximate formula for (3) is

$$N=\sigma_D^2(Z_\alpha+Z_\beta)^2/\Delta^2 \tag{5}$$

where σ_D is the standard deviation of the *paired-differences*. That is, each subject now has two measurements $x_i^{(1)}$ and $x_i^{(2)}$ on variable X at the two points in time. The quantity σ_D is the standard deviation of the paired differences $d_i=x_i^{(2)}-x_i^{(1)}$. The term σ in formula (3) refers to *between subject* variation, whereas the term σ_D in formula (5) refers to *within subject* variation. In general these quantities will be quite different. Sample size based on formula (5) corresponds to using a paired t-test for statistical analysis. Sample size based on formula (3) corresponds to using a two-sample t-test.

The sample size formula corresponding to (4) is somewhat more complicated for the cohort case. Its presentation and discussion is beyond the scope of this paper. If one had a cohort(s) observed at several points in time, one could consider sample size determinations based upon repeated measures designs. The book by Winer [10] provides a handy reference for this material.

The matter of calculating sample size to estimate group rates of change, or to detect differences in rates of change, becomes entwined with problems of frequency of measurement and study duration. These problems will be systematically treated in a subsequent paper.

4. COMMENT

There are a number of inherent limitations to these techniques, a few of which I wish to emphasize. That one must have some preliminary ideas about the magnitudes of the quantities to be estimated, or the differences to be detected, and their standard deviations, is no handicap. Few studies are undertaken in total ignorance, so that accumulated experience can be a guide in these matters. More serious is that no account has been taken of interrelationships among variables. More often than not, interrelationships among a constellation of variables will be unknown or poorly understood. A major goal of the study might be the discovery of interrelationships and the elucidation of their origin and mode of action. Although one could attempt to apply multiple regression and correlation techniques to handle these problems, the difficulties multiply rapidly. Presently my own position regarding the multivariate case is that unless one can be extraordinarily explicit about the forms of the relationships and the variables involved, one will have to be satisfied with simply doing a good job of planning for each variable, one at a time. This in itself can be a Herculean task, but if well done it should be a step in the right direction for the multivariate case.

Another matter is that the outlined techniques assume the data to have approximate normal distributions. A nonlinear transformation may improve the approximation to normality if the original form is markedly non-normal. See Kruskal's general review [11]. Over the years a number of papers have discussed approaches to sample size determination in the categorical case. One may consult the paper by Pasternack and Gilbert [12] as a recent example, along with its references. Cochran [13, 14] provides a more general discussion of problems in planning observational studies, and strategies in overcoming them. Baltes [15] and Schaie [16] consider various research designs.

One aspect of statistical theory and practice upon which I have not touched is sequential design, experimentation, and practice. Implicit in my discussion has been the assumption that a longitudinal study is primarily concerned with accumulating knowledge so as to characterize and elucidate biological phenomena, and that there are no ethical problems of potentially adverse effects of the study. If ethical considerations are important to the conduct and continuation of a study, sequential methods may be appropriate. Armitage [17] gives an authoritative survey of this field.

Acknowledgements—I wish to thank Dr. Max Halperin of the National Heart and Lung Institute for many enlightening conversations on this topic. Also appreciated are the stimulating discussions with Drs. Reubin Andres, Nathan Shock, and Mr. Arthur Norris of the Gerontology Research Center, Baltimore, Md.

REFERENCES

1. Shock NW, Yiengst MJ: Age changes in basal repiratory measurements and metabolism in males. **J Gerontol** 10: 31–40, 1955
2. Kish L: **Survey Sampling.** Wiley, New York, 1965
3. Cox DR: **Planning of Experiments.** Wiley, New York, 1958
4. Pearson ES, Hartley HO: **Biometrika Tables for Statisticians,** 3rd Edition. Cambridge University Press, London, 1966
5. Dixon WJ, Massy FJ: **Introduction to Statistical Analysis,** 3rd Edition. McGraw-Hill, New York, 1969
6. O'Neill R, Wetherill GB: The present state of multiple comparison methods (with Discussion). **J R Statist Soc** 33B: 218–250, 1971
7. Kastenbaum MA, Hoel DG, Bowman KO: Sample size requirements: one-way analysis of variance. **Biometrika** 57: 421–430, 1970
8. Kastenbaum MA, Hoel DG, Bowman KO: Sample size requirements: randomized block designs. **Biometrika** 57: 573–577, 1970
9. Bowman KO: Tables of the sample size requirement. **Biometrika** 59: 234, 1972
10. Winer BJ: **Statistical Principles in Experimental Design.** McGraw-Hill, New York, 1962
11. Kruskal JB: Transformations of data. **International Encyclopedia of the Social Sciences,** Vol. 15, pp. 182–193. Crowell-Collier and Free Press, New York, 1968
12. Pasternack BS ,Gilbert HS: Planning the duration of long-term survival time studies designed for accrual by cohorts. **J Chron Dis** 24: 681–700, 1971
13. Cochran WG: The planning of observational studies of human populations (with Discussion). **J R Statist Soc** 128A, 234–265, 1965
14. Cochran WG: Observational studies. **Statistical Paper in Honor of George W. Snedecor** (Ed. by T. A. Bancroft). Iowa State University Press, 1972
15. Baltes PB: Longitudinal and cross-sectional sequences in the study of age and generation effects. **Human Devel** 11: 145–171, 1968
16. Schaie KW: A general model for the study of developmental problems. **Psychol Bull** 64: 92–107, 1965
17. Armitage P: Some developments in the theory and practice of sequential medical trials. **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,** Vol. IV, pp. 791–804 (Ed. by L. Le Cam and J. Neyman). University of California Press, 1967

J Chron Dis 1973, Vol. 26, pp. 561-570.

PLANNING A LONGITUDINAL STUDY: II. FREQUENCY OF MEASUREMENT AND STUDY DURATION

JAMES J. SCHLESSELMAN

Biometry Branch,
National Institute of Child Health and Human Development,
Bethesda, Md. 20014, U.S.A.

1. INTRODUCTION

TWO IMPORTANT issues in planning and executing a longitudinal study are the frequency of repeated measurements on the subjects and the study duration. The scientific goals of a study, physical resources, available staff and funds will contribute to an investigator's intuition regarding the proper balance of these two factors. Statistical techniques can aid in planning and in evaluating choices of frequency of measurement and study duration. They assume that characterizing individual change in function with age (time) is an important component of a longitudinal study.

2. LINEAR CHANGE IN FUNCTION WITH AGE

One of the simplest descriptions of change in function with age is that provided by a straight line.

$$f=\alpha+\beta t.$$

The term f denotes the function value (e.g. creatinine clearance, basal metabolism, weight) at age t, and the parameters α and β denote respectively the intercept and slope of the line. The interpretation of β is that it represents the 'rate of aging' with regard to the function; α represents the function value at age zero. That any measured physiological function would perfectly relate to age in this simple fashion, even over a restricted range, is unlikely. However, one might be willing to admit that apart from fluctuations or 'errors' the relationship between function and age is linear. In this case one could write

$$f=\alpha+\beta t+\text{error}. \tag{1}$$

The notion is that corresponding to each value of t there is a set of potential values of f which depend upon the errors. One conceptualizes errors as having a distribution with mean zero and variance σ^2 at each value of t. The quantity σ^2 is a measure of variation *around the line*. If the errors are small (i.e. σ^2 is small), a straight line will be a good approximation or 'fit' in equation (1).

In a longitudinal study 'errors' depend upon a number of factors, among them: inherent biological variation over time in the measured organism, error of measurement in laboratory determinations over time, and departures of the 'true' relationship from linearity.

When given a sample of data on an individual at P points in time $(f_1, t_1), (f_2, t_2), \ldots, (f_p, t_p)$, one can form estimates of the intercept, slope, and error. One would not expect every individual to have the same rate of aging, so the αs and βs would be expected to vary among individuals. One therefore fits a line to the data on each individual. The sample estimates of α and β are denoted $\hat{a}$ and $\hat{b}$ respectively. Letting $\bar{f}=\Sigma f_i/P$ and $\bar{t}=\Sigma t_i/P$, the least squares estimates are given [1] by

$$\begin{aligned}\hat{b}&=\Sigma(f_i-\bar{f})(t_i-\bar{t})/\Sigma(t_i-\bar{t})^2\\ \hat{a}&=\bar{f}-\hat{b}\bar{t}.\end{aligned} \tag{2}$$

An estimate of σ^2 is given by

$$S^2=\{\Sigma(f_i-\bar{f})^2-\hat{b}\Sigma(f_i-\bar{f})(t_i-\bar{t})\}/(P-2). \tag{3}$$

The sample estimate S^2 is based upon the deviations from the fitted line. The quantities $\hat{a}$, $\hat{b}$ and S^2 are calculated separately for each subject. For convenience I am omitting the subscript 'k' which would appear on these terms for the kth subject.

The statistical techniques of this section are based upon providing 'good' estimates of individual slopes. If the relationship between a given function and age is linear, one can regard these slopes as being the rate of aging of an individual with regard to that function.

To gain an appreciation of how the number of measurements and study duration affect the estimate of an individual's slope, consider the following set-up. Suppose one has measurements at P equally spaced points in time, $t_1, t_2, \ldots, t_p$, where each point is spaced Y years apart. The study duration D is given by

$$D=Y(P-1) \tag{4}$$

and the frequency of measurement f is simply

$$f=(P-1)/D. \tag{5}$$

The standard error of $\hat{b}$ is given [1] by S.E. $(\hat{b})=\sigma/\sqrt{\Sigma(t_i-\bar{t})^2}$. Using well-known induction formulas (e.g. [2] p. 31, exercises 5 and 6) one can easily show that $\Sigma(t_i-\bar{t})^2=Y^2P(P-1)(P+1)/12=D^2P(P+1)/[12(P-1)]$. The standard error of $\hat{b}$ can therefore be written as

$$\text{S.E. }(\hat{b})=\sigma\sqrt{12(P-1)}/[D\sqrt{P(P+1)}]. \tag{6}$$

Equation (6) shows explicitly how the standard error of $\hat{b}$ depends on the study duration D and the number of repeated measurements P on a subject. Since we have approximately that S.E. $(\hat{b})\propto 1/[D\sqrt{P}]$, a 'unit' increase in D reduces the standard error more than a unit increase in P.

Tables 1–3 are presented to facilitate the process of evaluating choices of study duration and number of measurements. These tables show values of

$$\omega=\sqrt{12(P-1)}/[D\sqrt{P(P+1)}] \tag{7}$$

for $P=2$–30 points and for studies of duration $D=1$–30 yr. Since

$$\mathrm{S.E.}(\hat{b}) = \omega\sigma \tag{8}$$

one can regard the ω values as indices of precision of $\hat{b}$. The smaller the value of ω, the smaller will be S.E. ($\hat{b}$) and, consequently, the more precise will be our estimate $\hat{b}$. (The first two lines of Table 1 are identical because of rounding in the third decimal place.)

TABLE 1. TABLED BY DURATION OF STUDY IN YEARS D AND NUMBER OF REPEATED OBSERVATIONS P ARE VALUES OF $\omega = \sqrt{12(P-1)}/[D\sqrt{P(P+1)}]$

	D									
P	1	2	3	4	5	6	7	8	9	10
2	1.414	0.707	0.471	0.354	0.283	0.236	0.202	0.177	0.157	0.141
3	1.414	0.707	0.471	0.354	0.283	0.236	0.202	0.177	0.157	0.141
4	1.342	0.671	0.447	0.335	0.268	0.224	0.192	0.168	0.149	0.134
5	1.265	0.632	0.422	0.316	0.253	0.211	0.181	0.158	0.141	0.126
6	1.195	0.598	0.398	0.299	0.239	0.199	0.171	0.149	0.133	0.120
7	1.134	0.567	0.378	0.283	0.227	0.189	0.162	0.142	0.126	0.113
8	1.080	0.540	0.360	0.270	0.216	0.180	0.154	0.135	0.120	0.108
9	1.033	0.516	0.344	0.258	0.207	0.172	0.148	0.129	0.115	0.103
10	0.991	0.495	0.330	0.248	0.198	0.165	0.142	0.124	0.110	0.099
11	0.953	0.477	0.318	0.238	0.191	0.159	0.136	0.119	0.106	0.095
12	0.920	0.460	0.307	0.230	0.184	0.153	0.131	0.115	0.102	0.092
13	0.889	0.445	0.296	0.222	0.178	0.148	0.127	0.111	0.099	0.089
14	0.862	0.431	0.287	0.215	0.172	0.144	0.123	0.108	0.096	0.086
15	0.837	0.418	0.279	0.209	0.167	0.139	0.120	0.105	0.093	0.084
16	0.813	0.407	0.271	0.203	0.163	0.136	0.116	0.102	0.090	0.081
17	0.792	0.396	0.264	0.198	0.158	0.132	0.113	0.099	0.088	0.079
18	0.772	0.386	0.257	0.193	0.154	0.129	0.110	0.097	0.086	0.077
19	0.754	0.377	0.251	0.188	0.151	0.126	0.108	0.094	0.084	0.075
20	0.737	0.368	0.246	0.184	0.147	0.123	0.105	0.092	0.082	0.074
21	0.721	0.360	0.240	0.180	0.144	0.120	0.103	0.090	0.080	0.072
22	0.706	0.353	0.235	0.176	0.141	0.118	0.101	0.088	0.078	0.071
23	0.692	0.346	0.231	0.173	0.138	0.115	0.099	0.086	0.077	0.069
24	0.678	0.339	0.226	0.170	0.136	0.113	0.097	0.085	0.075	0.068
25	0.666	0.333	0.222	0.166	0.133	0.111	0.095	0.083	0.074	0.067
26	0.654	0.327	0.218	0.163	0.131	0.109	0.093	0.082	0.073	0.065
27	0.642	0.321	0.214	0.161	0.128	0.107	0.092	0.080	0.071	0.064
28	0.632	0.316	0.211	0.158	0.126	0.105	0.090	0.079	0.070	0.063
29	0.621	0.311	0.207	0.155	0.124	0.104	0.089	0.078	0.069	0.062
30	0.612	0.306	0.204	0.153	0.122	0.102	0.087	0.076	0.068	0.061

As an example, suppose one wants to consider a 10 yr study. By reading down column 10 in Table 1, we find that for 2 points, $\omega=0.141$; for 5 points, $\omega=0.126$; for 10 points, $\omega=0.099$; for 30 points, $\omega=0.061$. Thus for a study of fixed length, one can easily find how increasing the number of observations during the study time increases the precision of the estimate of a slope. Using formula (4) one finds the spacing between observations. For example, for $P=2$ points in $D=10$ yr, $Y=10$. Thus observations are made at times 0 and 10. For $P=5$ points in $D=10$ yr, $Y=2.5$. Thus observations are made at times 0, 2.5, 5, 7.5 and 10. From formula (5), the frequency of measurement for these two examples is $f=1/10$, once every 10 yr and $f=4/10$, once every 2.5 yr.

As another example, suppose one wishes to measure each subject only twice. By looking across the row for $P=2$ in Tables 1–3, one can easily find how increasing the

TABLE 2. TABLED BY DURATION OF STUDY IN YEARS D AND NUMBER OF REPEATED OBSERVATIONS P ARE VALUES OF $\omega=\sqrt{12(P-1)}/[D\sqrt{P(P+1)}]$

	D									
P	11	12	13	14	15	16	17	18	19	20
2	0.129	0.118	0.109	0.101	0.094	0.088	0.083	0.079	0.074	0.071
3	0.129	0.118	0.109	0.101	0.094	0.088	0.083	0.079	0.074	0.071
4	0.122	0.112	0.103	0.096	0.089	0.084	0.079	0.075	0.071	0.067
5	0.115	0.105	0.097	0.090	0.084	0.079	0.074	0.070	0.067	0.063
6	0.109	0.100	0.092	0.085	0.080	0.075	0.070	0.066	0.063	0.060
7	0.103	0.094	0.087	0.081	0.076	0.071	0.067	0.063	0.060	0.057
8	0.098	0.090	0.083	0.077	0.072	0.068	0.064	0.060	0.057	0.054
9	0.094	0.086	0.079	0.074	0.069	0.065	0.061	0.057	0.054	0.052
10	0.090	0.083	0.076	0.071	0.066	0.062	0.058	0.055	0.052	0.050
11	0.087	0.079	0.073	0.068	0.064	0.060	0.056	0.053	0.050	0.048
12	0.084	0.077	0.071	0.066	0.061	0.057	0.054	0.051	0.048	0.046
13	0.081	0.074	0.068	0.064	0.059	0.056	0.052	0.049	0.047	0.044
14	0.078	0.072	0.066	0.062	0.057	0.054	0.051	0.048	0.045	0.043
15	0.076	0.070	0.064	0.060	0.056	0.052	0.049	0.046	0.044	0.042
16	0.074	0.068	0.063	0.058	0.054	0.051	0.048	0.045	0.043	0.041
17	0.072	0.066	0.061	0.057	0.053	0.050	0.047	0.044	0.042	0.040
18	0.070	0.064	0.059	0.055	0.051	0.048	0.045	0.043	0.041	0.039
19	0.069	0.063	0.058	0.054	0.050	0.047	0.044	0.042	0.040	0.038
20	0.067	0.061	0.057	0.053	0.049	0.046	0.043	0.041	0.039	0.037
21	0.066	0.060	0.055	0.051	0.048	0.045	0.042	0.040	0.038	0.036
22	0.064	0.059	0.054	0.050	0.047	0.044	0.042	0.039	0.037	0.035
23	0.063	0.058	0.053	0.049	0.046	0.043	0.041	0.038	0.036	0.035
24	0.062	0.057	0.052	0.048	0.045	0.042	0.040	0.038	0.036	0.034
25	0.061	0.055	0.051	0.048	0.044	0.042	0.039	0.037	0.035	0.033
26	0.059	0.054	0.050	0.047	0.044	0.041	0.038	0.036	0.034	0.033
27	0.058	0.054	0.049	0.046	0.043	0.040	0.038	0.036	0.034	0.032
28	0.057	0.053	0.049	0.045	0.042	0.039	0.037	0.035	0.033	0.032
29	0.056	0.052	0.048	0.044	0.041	0.039	0.037	0.035	0.033	0.031
30	0.056	0.051	0.047	0.044	0.041	0.038	0.036	0.034	0.032	0.031

time span over which these 2 measurements are made decreases S.E. ($\hat{b}$). Thus for $D=1$, $\omega=1.414$; for $D=5$, $\omega=0.283$; for $D=15$, $\omega=0.094$; for $D=30$, $\omega=0.047$.

A more useful application of Tables 1–3 comes from considering the isoclines of ω. That is, consider all combinations of P and D for which ω is constant. These provide alternative 'strategies' which yield equal precision for S.E. ($\hat{b}$). For example, the pairs (P, D) corresponding to $\omega=0.20$ are (3, 7), (6, 6), (9, 5), (17, 4) and (30, 3). One can interpolate in Tables 1–3 for studies with fractional lengths of years, or use formula (7) directly. This example shows that using 3 points over 7 yr gives just as precise an estimate as using 30 points over 3 yr.

I wish to emphasize that an inherent limitation in the application of this technique is the assumption of linearity upon which it rests. For this reason one should be wary of strategies that recommend many measurements over a short period of time, or few measurements over a long period of time. In the preceding example, the strategies (17, 4) and (30, 3) might not be wise in practice. Though this technique assumes linearity as a simplification for planning, one should be prepared to encounter nonlinearities in one's data, and to arrange statistical analyses to detect them. That we assume linearity for planning convenience, but find that our data is nonlinear, does not invalidate the statistical analysis of the data or the conclusions based upon it. The simplest of nonlinearities is quadratic curvature, a brief discussion of which appears in Section 5.

TABLE 3. TABLED BY DURATION OF STUDY IN YEARS D AND NUMBER OF REPEATED OBSERVATIONS P ARE VALUES OF $\omega = \sqrt{12(P-1)}/[D\sqrt{P(P+1)}]$

	D									
P	21	22	23	24	25	26	27	28	29	30
2	0.067	0.064	0.061	0.059	0.057	0.054	0.052	0.051	0.049	0.047
3	0.067	0.064	0.061	0.059	0.057	0.054	0.052	0.051	0.049	0.047
4	0.064	0.061	0.058	0.056	0.054	0.052	0.050	0.048	0.046	0.045
5	0.060	0.057	0.055	0.053	0.051	0.049	0.047	0.045	0.044	0.042
6	0.057	0.054	0.052	0.050	0.048	0.046	0.044	0.043	0.041	0.040
7	0.054	0.052	0.049	0.047	0.045	0.044	0.042	0.040	0.039	0.038
8	0.051	0.049	0.047	0.045	0.043	0.042	0.040	0.039	0.037	0.036
9	0.049	0.047	0.045	0.043	0.041	0.040	0.038	0.037	0.036	0.034
10	0.047	0.045	0.043	0.041	0.040	0.038	0.037	0.035	0.034	0.033
11	0.045	0.043	0.041	0.040	0.038	0.037	0.035	0.034	0.033	0.032
12	0.044	0.042	0.040	0.038	0.037	0.035	0.034	0.033	0.032	0.031
13	0.042	0.040	0.039	0.037	0.036	0.034	0.033	0.032	0.031	0.030
14	0.041	0.039	0.037	0.036	0.034	0.033	0.032	0.031	0.030	0.029
15	0.040	0.038	0.036	0.035	0.033	0.032	0.031	0.030	0.029	0.028
16	0.039	0.037	0.035	0.034	0.033	0.031	0.030	0.029	0.028	0.027
17	0.038	0.036	0.034	0.033	0.032	0.030	0.029	0.028	0.027	0.026
18	0.037	0.035	0.034	0.032	0.031	0.030	0.029	0.028	0.027	0.026
19	0.036	0.034	0.033	0.031	0.030	0.029	0.028	0.027	0.026	0.025
20	0.035	0.033	0.032	0.031	0.029	0.028	0.027	0.026	0.025	0.025
21	0.034	0.033	0.031	0.030	0.029	0.028	0.027	0.026	0.025	0.024
22	0.034	0.032	0.031	0.029	0.028	0.027	0.026	0.025	0.024	0.024
23	0.033	0.031	0.030	0.029	0.028	0.027	0.026	0.025	0.024	0.023
24	0.032	0.031	0.029	0.028	0.027	0.026	0.025	0.024	0.023	0.023
25	0.032	0.030	0.029	0.028	0.027	0.026	0.025	0.024	0.023	0.022
26	0.031	0.030	0.028	0.027	0.026	0.025	0.024	0.023	0.023	0.022
27	0.031	0.029	0.028	0.027	0.026	0.025	0.024	0.023	0.022	0.021
28	0.030	0.029	0.027	0.026	0.025	0.024	0.023	0.023	0.022	0.021
29	0.030	0.028	0.027	0.026	0.025	0.024	0.023	0.022	0.021	0.021
30	0.029	0.028	0.027	0.025	0.024	0.024	0.023	0.022	0.021	0.020

3. EVALUATING STUDY DURATION AND NUMBER OF MEASUREMENTS

Two alternative methods of evaluating study duration and the number of repeated measurements involve using the standard error of $\hat{b}$ and the coefficient of variation of $\hat{b}$. If one requires that the standard error of $\hat{b}$ be equal to a specified quantity ε, then S.E. $(\hat{b}) = \varepsilon$ implies that

$$\omega = \varepsilon/\sigma \tag{9}$$

using equation (8). If one wishes instead to determine $\hat{b}$ with a specified coefficient of variation CV, then $CV(\hat{b}) = \text{S.E. } (\hat{b})/\hat{b}$ implies that

$$\omega = \hat{b}\ \text{CV}/\sigma. \tag{10}$$

To use formulas (9) and (10) for planning purposes, one must specify *anticipated* values for $\hat{b}$ and σ and specify desired values for ε or CV. Substituting these values into equation (9) or (10) gives a value for ω. By then referring to Tables 1–3, one can find which strategies of (P, D) yield the calculated ω. (In using formula (10) one should ignore the sign of $\hat{b}$).

In Section 2 I stated that each individual would have his own S^2 and $\hat{b}$. One way of using formulas (9) and (10), but by no means the only way, is to use some *average* value for $\hat{b}$ and some average value for σ. Thus in a sense we would be planning on the basis of some 'average person'. If one had longitudinal data at hand, he could use this data for planning purposes. Most likely only cross-sectional data will be available. Therefore the application of these techniques using available cross-sectional data will be discussed.

As an example, I will use data reported by Shock and Yiengst [3]. In Table 4 on p. 36, they give values of slopes based on cross-sectional age regressions for a number of variables. As two instances, they report that $b=-0.137$ for heat production and $b=-0.476$ for oxygen uptake. The value $b=-0.137$ means that on the average there is a cross-sectional decline in heat production of 0.137 cal/sq m/hr per yr. The value $b=-0.476$ means that on the average there is a cross-sectional decline in oxygen uptake of 0.476 cc/sq m/min per yr. Referring again to their Table 4, please note that the numbers in the column headed σ_b are *not* estimates of σ. The values of σ_b reported are the *standard errors of b based on cross-sectional data.* The quantity σ in this paper is the *standard deviation about the regression line within a single person.* Values of σ_b confound variation among persons with variation within persons. Furthermore σ_b has incorporated into it the factor $\Sigma(a_i-\bar{a})^2$, where $a_1, a_2, \ldots a_N$ are the ages of the N subjects at which the cross-sectional measurements were made. I am emphasizing these points to caution readers against using reported numbers based upon similarity of notation used. One must understand the derivation of the reported values and their meaning.

For estimates of σ^2, refer to Shock and Yiengst's Table 3 on p. 34 and look down the last column headed 'Error—608 df.' These values represent the average variation within subjects, and incorporate laboratory error of measurement and period-to-period variation. If the periods were widely separated in time, these values would better estimate σ^2. Because of the relatively close spacing of the periods in time, the reported error mean squares probably underestimate σ^2. Nevertheless they are useful, and with caution can be used as anticipated values in formulas (9) and (10). For oxygen uptake the reported error mean square of 41.17 gives an anticipated value of $\sigma=\sqrt{41.17}=6.42$. For heat production the reported error mean square of 2.87 gives an anticipated value of $\sigma=\sqrt{2.87}=1.69$.

If one assumes that longitudinal change within an individual will follow the observed cross-sectional change, then the reported values of b based upon cross-sectional analysis can be used as anticipated values for $\hat{b}$ in formula (10). There are many good reasons to be cautious about this assumption. See for example the work of Baltes [4] and Schaie and Strother [5].

Assume for the moment that longitudinal change in heat production will follow the cross-sectional gradient, that $\hat{b}=-0.137$ and that $\sigma=1.69$. Furthermore suppose that one wants to determine individual slopes in heat production with an average *CV* of 50 per cent. Using formula (10) and ignoring the sign of $\hat{b}$, we have $\omega=(0.137)(0.50)/1.69 =0.041$. Referring to Tables 1–3, one can find those strategies for which $\omega=0.041$. Writing the number of measurements P and the study duration D as the ordered pair (P, D), some strategies for which $\omega=0.041$ are: (14, 21), (11, 23), (10, 24), (9, 25) and (6, 29). Thus studies ranging from 14 points in 21 yr to 6 points in 29 yr will yield slopes on heat production with an average coefficient of variation of 50 per cent.

One can pursue this exercise using formulas (9) or (10) for any variable to be studied. One will be confronted no doubt with a bewildering array of alternative strategies.

Some compromise must be reached, most likely based upon considerations mentioned in Section 1. Having decided upon a strategy, one can then use formulas (6)–(10) with the chosen values of P and D to estimate the precision anticipated from the study. The actual precision may differ from the planned precision because we are using 'guesses' for $\hat{b}$ and σ. One advantage of using formula (9) instead of (10) is that one need only provide an anticipated value for σ. An underestimate of σ on any variable is given by the standard deviation of replicate laboratory determinations. This underestimates σ because it does not incorporate the additional factors of inherent biological variation in the organism over time and departures of the 'true' relationship from linearity. For most variables the inherent biological variation may be manyfold greater than laboratory measurement error. Circadian rhythms in hormone secretion, or change in body function under the influence of temperature, light, season, diet or infection [6] are but some examples of what I mean by inherent biological variation. Elimination of these potential sources of variability through careful experimental design can yield great increases in precision.

For additional information, including advice on using *CV*, refer to [7]. Please observe that the symbols σ and P have different meanings in these two papers.

4. ADDITIONAL COMPLICATIONS

The matter of study duration and number of repeated measurements has been treated without referring to the sample size. This occurred because the focus was on obtaining good estimates of slope for an ('average') individual. The problem of sample size arises when one wants to estimate the average rate of change for various groups, or when one wants to detect differences among groups in their average rates of change. The groups might be determined, for example, by birth-cohorts or by classification on clinical or social variables.

Consider a group of N persons, each having a 'true' rate of change $\beta_1, \beta_2, \ldots, \beta_N$. If one measures each person at P points in time, he can calculate for each person an estimate of slope $\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_N$ using formula (2). The true average rate of change for the group is $\bar{\beta} = \{\beta_1 + \ldots + \beta_N\}/N$. An estimate of $\bar{\beta}$ is $\bar{b} = \{\hat{b}_1 + \ldots \hat{b}_N\}/N$. The standard error of $\bar{b}$ is given by

$$\text{S.E. } (\bar{b}) = \{\sigma_\beta^2 + \sigma^2/\Sigma(t_i - \bar{t})^2\}^{\frac{1}{2}}/\sqrt{N}. \tag{11}$$

The quantity σ_β^2 represents the variation in true slopes among people, where $\sigma_\beta^2 \approx \Sigma(\beta_i - \bar{\beta})^2/N$. The quantity σ^2 has the same meaning as in Sections 2 and 3, namely, the variation about the regression line within a person. For simplicity I am assuming that σ^2 does not vary from person to person. To avoid additional complications in exposition, I also assume the approximate equality $\sigma_\beta^2 \approx \Sigma(\beta_i - \bar{\beta})^2/N$.

Using the results in Section 2, one can rewrite equation (11) as

$$\text{S.E. } (\bar{b}) = \{\sigma_\beta^2 + 12(P-1)\sigma^2/[D^2P(P+1)]\}^{\frac{1}{2}}/\sqrt{N}. \tag{12}$$

Equation (12) shows that one can reduce S.E. $(\bar{b})$ in three different ways: (i) increase the sample size, N; (ii) increase the number of points, P; (iii) increase the study duration, D. Which of these alternatives is most profitable, or in what measure one should combine them, is a very difficult problem in practice. The answer will depend upon the relative sizes of the variance components σ_β^2 and σ^2, and on the relative costs of unit increases in N, P and D. Note in equation (12) that all three factors N, P and D contribute to decreasing the term in σ^2, whereas only N acts to decrease the term in σ_β^2.

One needs a theoretical model or longitudinal data to get estimates or anticipated values for σ_β^2. If one had anticipated values for σ_β^2 and σ^2, these could be substituted into equation (12). One could then use equations (1), (2) or (3) in Ref. [7] to obtain simultaneously values for N, P and D. This would be somewhat complicated because of the multiplicity of solution sets (N, P, D). For example, suppose one wants to estimate the average rate of change within a group so that the standard error is equal to a specified quantity ε. Using equation (12) we have

$$\{\sigma_\beta^2+12(P-1)\sigma^2/[D^2P(P+1)]\}/N=\varepsilon^2. \tag{13}$$

Using anticipated values for σ_β^2 and σ^2, we now search for all values of (N, P, D) which solve equation (13). I would approach this task by, say, fixing P at several values, and then solving for N and D numerically. One may find it instructive to compare equation (13) with equation (1) in Ref. [7].

As another example, suppose one wanted to detect a difference of Δ units between the average rates of change in two groups. Using equation (3) in Ref. [7] we have

$$\{\sigma_\beta^2+12(P-1)\sigma^2/[D^2P(P+1)]\}/N=\Delta^2/[2(Z_\alpha+Z_\beta)^2]. \tag{14}$$

The quantities Z_α and Z_β are unit normal deviates corresponding to the desired Type I error rate α and the Type II error rate β and are given in Table 2 in [7]. One has to solve equation (14) numerically, just as one does (13). Considerations of Type I and Type II errors arise only in the context of *testing* for differences, that is, in testing hypotheses. These considerations do not apply to the development in Sections 2 and 3 which is based upon *estimation*.

The purpose of this section is to show that problems of sample size determination can become entwined with problems of study duration and frequency of measurement. Their separation is somewhat artificial, but may nevertheless be preferable for simplicity of analysis. Apart from speculating about the size of σ_β^2, one really needs longitudinal data at hand to use the techniques of this section. This needed data should begin to appear for many physiological variables in forthcoming publications of the Gerontology Research Center, Baltimore, Md. In this longitudinal study average values of N, P and D presently are: $N\approx 1000$, $P\approx 5$, $D\approx 10$. These are averages because of continuing recruitment and dropouts.

5. QUADRATIC CHANGE IN FUNCTION WITH AGE

One does not expect all functional relationships with age to be linear. Planning that accounts only for linear trends should be adequate, however. Basing statistical analyses solely on linear trends may be satisfactory as a first approximation, but it will be insufficient for a more refined analysis. One of the simplest functions allowing for curvature is a 2nd degree polynomial

$$f=\beta_0+\beta_1t+\beta_2t^2. \tag{15}$$

Figure 1 displays some shapes of the function (15) for various values of the parameters. The parameter β_0 represents the function value at $t=0$. The parameter β_2 denotes the 'accelerative' component of aging. This is because the 2nd derivative of f with regard

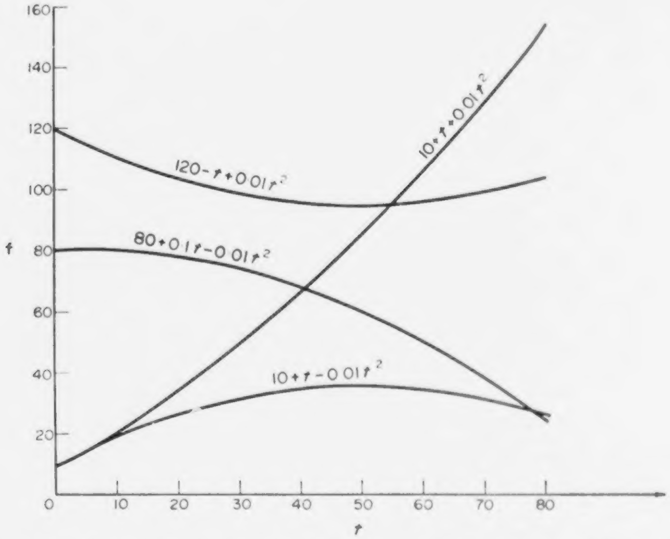


FIG. 1. The function $f=\beta_0+\beta_1 t+\beta_2 t^2$ is graphed for various values of the parameters β_0, β_1 and β_2.

to t is $d^2f/dt^2=2\beta_2$. The parameter β_1 represents the rate of aging at $t=0$. This is because

$$df/dt=\beta_1+2\beta_2 t. \tag{16}$$

If $\beta_2=0$, as in a straight line, the rate of aging at $t=0$ is the same at every age. When $\beta_2\neq 0$, the rate of aging itself depends upon age according to equation (16). To allow for fluctuations or errors in an exact fit by (15), it is preferable to write

$$f=\beta_0+\beta_1 t+\beta_2 t^2+\text{error}.$$

As in Section 2, one would fit separate curves to each individual's data. Consult Snedecor and Cochran [8] for details of the fitting procedure.

The main distinction I wish to draw in this section is between techniques used for planning and those used for analysis. Whereas one may wish to restrict himself to simple functions like straight lines for purposes of planning, he should not limit himself in the analysis. The planning techniques of Section 2 could be easily extended to 2nd degree polynomials or other more complicated curves. My feeling is that the additional complications introduced into the planning process will not be worth the effort. The major burden, moreover, is in sensibly specifying desired values for ε, CV, or Δ, and in providing good anticipated values for $\hat{b}$, σ^2 and σ^2_β. The most sophisticated technique based upon faulty data serves no one well.

Acknowledgements—I wish to thank Dr. Max Halperin of the National Heart and Lung Institute for many enlightening conversations. Also appreciated are the many stimulating discussions with Drs. Reubin Andres, Nathan Shock, and Mr. Arthur Norris of the Gerontology Research Center, Baltimore, Md.

REFERENCES

1. Armitage P: **Statistical Methods in Medical Research.** Wiley, New York, 1971
2. Apostol TM: **Calculus.** Vol. I. Blaisdell, New York, 1961
3. Shock NW, Yiengst MJ: Age changes in basal respiratory measurements and metabolism in males. **J Gerontol** 10: 31–40, 1955
4. Baltes PB: Longitudinal and cross-sectional sequences in the study of age and generation effects. **Human Devel** 11: 145–171, 1968
5. Schaie KW, Strother CR: A cross-sequential study of age changes in cognitive behaviour. **Psychol Bull** 70: 671–680, 1968
6. Dubos R: **Man Adapting.** Yale University Press, New Haven, 1964
7. Schlesselman JJ: Planning a longitudinal study: I. Sample size determination. **J Chron Dis** 26: 553–560, 1973
8. Snedecor GW, Cochran WG: **Statistical Methods.** 6th Ed, Iowa State University Press, Ames, 1967

Carson, L.A. (Ed.),
Nutrition in Old Age. X Symposium Swedish Nutrition Foundation.
Uppsala: Almqvist & Wiksell, 1972

ENERGY METABOLISM, CALORIC INTAKE AND PHYSICAL ACTIVITY OF THE AGING

by N. W. Shock[1]

The central biological fact about aging is that mortality rates or the probability of death in adults increases with age. In the United States, mortality rate or probability of death from all causes for males increases from about 11 per 1 000 population at age 25–30 to 460 per 1 000 at age 85–90 (U.S. Department of Health, Education and Welfare, 1971). As indicated in Fig. 1, mortality rates increase rapidly as age advances and in fact double every $8\,^{1}/_{2}$ years. On the basis of these observations, and others, it can be assumed that with the passage of time changes occur in individuals which increase their vulnerability to a broad spectrum of conditions that result in death.

One of the goals of the Gerontology Research Center has been to describe in quantitative terms the physiological changes that occur in human subjects with advancing age. Ideally, measurements should be made in the same individual as he ages, if we are to identify age changes. This longitudinal method is laborious and time consuming so that although we have been conducting such a study on about 650 normal males for the past 10 years, most of our current information is based on cross sectional analyses of age differences. In these studies mean values by decades are calculated for different subjects, or the regression of measurements for a given variable on chronological age are calculated.

Unless specified, the observations presented here represent cross sectional analyses of data obtained from normal subjects between the ages of 20 and 90 years. All subjects were carefully examined by physicians on the staff of the Gerontology Research Center and any subjects with identifiable disease of the organ systems involved in the measurements were excluded from the data analysis. Since we regard aging as a process which goes on over the entire adult life span, observations were made on subjects (males) between the ages of 20 and 90 years.

Before turning to the specific topics of energy metabolism, caloric intake, and physical activity, I shall present a brief summary of our results with respect to the general physiological characteristics of aging in humans.

In the first place, we have found that a number of characteristics of the blood, measured under resting or basal conditions are well maintained even in subjects of advanced age. Fasting blood sugar levels (Smith and Shock, 1949), pH of the plasma (Shock and Yiengst, 1950), plasma volume (Cohn and Shock, 1949) and osmotic pressure are examples of characteristics which, under resting or basal conditions show little or no change with age. As will be shown later, the rate of recovery to resting values

---

[1] Gerontology Research Center, National Institute of Child Health and Human Development, National Institutes of Health, Baltimore City Hospitals, Baltimore, Maryland 21224, USA.

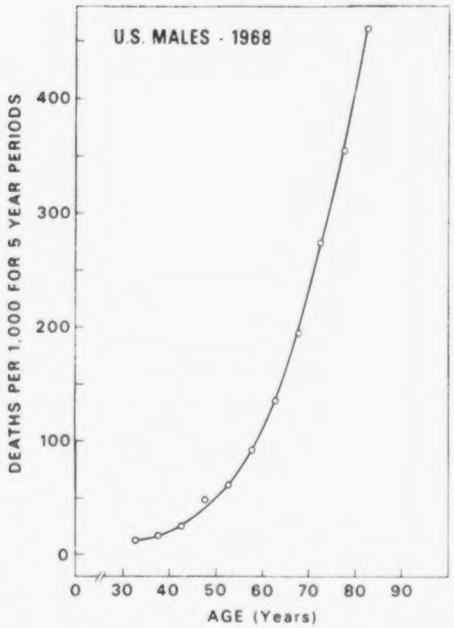


Fig. 1. Deaths per 1 000 for 5 year periods. United States males, 1968. (U.S. Dept. Health, Education and Welfare, 1971.)

following physiological displacements, may be slower in old subjects than in young.

A good many physiological functions show a gradual decrement with increasing age. Resting cardiac output, measured by the dye dilution technique, falls progressively with increasing age (Brandfonbrener, Landowne and Shock, 1955). Average values for cardiac index at age 80 are only about 30% of the average values for 30-year-olds. Renal plasma flow, measured by the clearance of diodrast or para-amino hippuric acid declines by about 50% between the ages of 30 and 80 years (Davies and Shock, 1950; Shock, 1961) (Fig. 2). Other functions, such as maximum breathing capacity and maximum oxygen uptake during physical exercise decline as much as 60–70% over the age span of 30 to 80 years (Norris, Shock and Yiengst, 1955; Norris and Shock, 1971). Fig. 3 illustrates the decrement in muscle strength, measured with a hand dynamometer in males and females over the age span 30 to 90 years. Maximum strength for the dominant hand in males fell from about 43 kg at age 30 to 25 kg at age 90 (Miles, 1950).

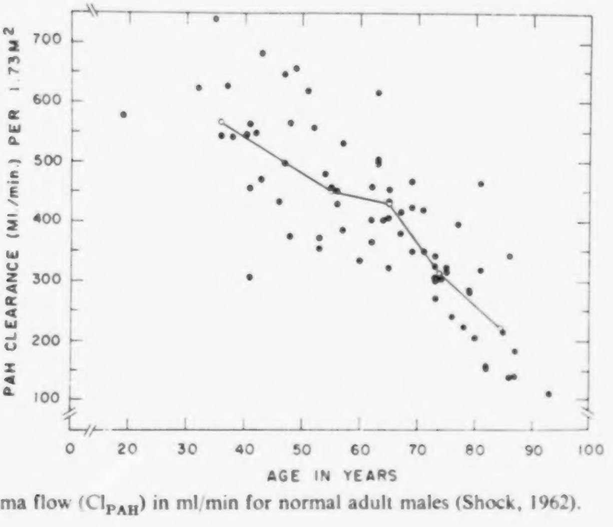


Fig. 2. Renal plasma flow (Cl_{PAH}) in ml/min for normal adult males (Shock, 1962).

Although aging is associated with a decline in average values for many physiological functions, it is important to recognize the wide range of values seen among different subjects at each decade of age. Fig. 2 shows that in one 80-year-old subject renal blood flow was still as good as that of the average 50-year-old. The effects of age are highly individual, and chronological age alone is a poor index of physiological function. Similar individual differences in the effect of age appear in every physiological system that we have measured.

When extra demands are imposed on an organ system, age differences are more pronounced than when observations are made under resting or basal conditions. The old individual shows a greater displacement and slower rate of recovery than does the young. For example, the oral administration of 10 grams of ammonium chloride to a young subject will produce a reduction in the plasma pH of approximately 0.05 units in 1 $^1/_2$ hours with complete recovery to the resting level by the end of 8–10 hours. However, in the 80-year-old the same dose will produce a displacement of about three

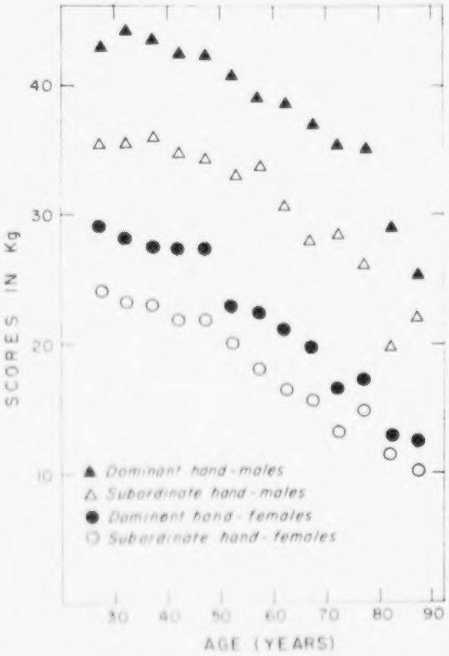


Fig. 3. Age decrement in muscle strength. ▲, Dominant hand, males; △, Subordinate hand, males; ●, Dominant hand, females; ○, Subordinate hand, females. (Miles, 1950.)

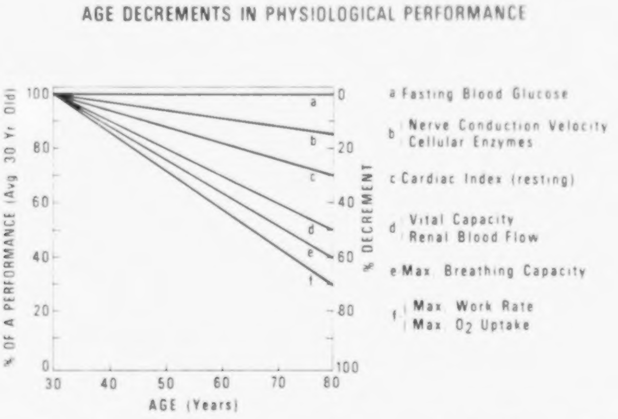


Fig. 4. Age decrements in physiological performances. Average values for 30-year-old subjects taken as 100%. Decrements shown are schematic. (*a*) Fasting blood glucose, (*b*) Nerve conduction velocity and some cellular enzyme activities, (*c*) Resting cardiac index, (*d*) Vital capacity and renal blood flow, (*e*) Maximum breathing capacity, (*f*) Maximum work rate and maximum O_2 uptake.

times this amount and recovery will require from 24 to 72 hours (Shock, 1961). The aged individual is able to make the adjustment but he requires more time. Similar impairments in the rate of recovery of the elderly following displacing stimuli have also been shown for the rate of removal of excess glucose from the blood (Silverstone, Brandfonbrener, Shock and Yiengst, 1957), the recovery of respiratory and cardiovascular displacements induced by standardized light exercise (Norris, Shock and Yiengst, 1953) and many other functions. This reduced speed of response is a general characteristic of aging which extends from physiological to psychological characteristics (Hügin, Norris and Shock, 1960; Suci, Davidoff and Surwillo, 1960).

The effects of age differ widely among organ systems. Fig. 4 illustrates the range of these differences. In this figure the average decrement between the ages of 30 and 80 years is plotted as the percentage of the mean value observed in 30-year-olds. Average decrements range from 15–20% for nerve conduction velocity (Norris, Shock and Wagman, 1953) to 70% for maximum oxygen uptake during maximum exercise. It can be seen that the greatest age decrements are found in tests which impose a stress on the organism and require the coordinated activity of a number of organ systems—as for example, physical exercise.

Part of the loss in reserve capacity observed in the elderly can be ascribed to the loss of functional elements in some organs. For example, the decline in kidney function is associated with a gradual loss of functioning nephrons (Arataki, 1926). In the senescent rat, degeneration of muscle fibers occurs (Andrew, Shock, Barrows and Yiengst, 1959) and the number of fibers in nerve trunks diminishes (Rexed, 1944). Although there are species differences in the extent to which cells or functioning elements are lost from different organ systems (Buetow, 1971), the loss in reserve capacities in the elderly must be in part due to tissue losses and cell death.

There is, however, an expanding body of data which indicates that with aging, changes also occur in the biochemical and physiological characteristics of tissues and cells which impair their function. Time will not permit a detailed review of this important aspect of aging.

It has long been known that the total energy production, per 24 hours or per sq.m. of surface area falls progressively with age (Boothby, Berkson and Dunn, 1936). Other studies have repeatedly confirmed an average decrement of about 12 cals/M^2/hr between the ages of 20 and 90 (Shock, 1955). Although the absolute levels vary by about 2 cal/M^2/hr, the slopes of the regression on age (1.66 cal/M^2/decade) are remarkably uniform except for one study (Lewis, 1938) in which the regression slope was less (0.8 cal/M^2/decade). This study was conducted on normal subjects from New York City who came to the hospital for tests on an out patient basis, whereas the other studies were made on subjects living in an institution.

These results led to the assumption that aging was associated with a "slowing of metabolism", presumably at the cellular level. Since thyroid hormone is the primary regulator of the rate of cellular metabolism, studies were directed toward determining age changes in the ability of the gland to produce or release thyroxine.

Estimated as the plasma protein bound iodine (PBI) the circulating thyroid hormone is unaltered throughout the adult age span (Gaffney, Gregerman and Shock, 1960). This finding means only that the gland is capable of maintaining a steady state of

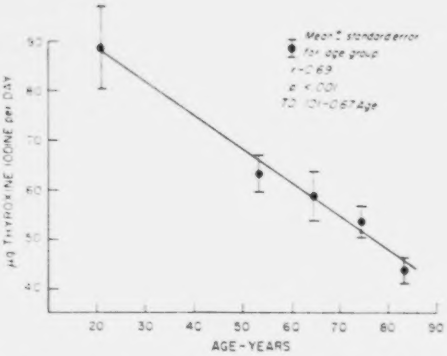


Fig. 5. μg thyroxine degraded per day in males aged 20–85 years (Gregerman, Gaffney and Shock, 1962).

PBI in the plasma. It does not give any indication of the rate of formation of the hormone.

Evidence on this question is offered by determining the rate of disappearance from the blood of radioactive thyroxine labeled with I^{131} (Gregerman, Gaffney and Shock, 1962). In this technique a measured amount of the labeled hormone was administered intravenously and its disappearance from the circulation was followed over a 2 week period. The daily turnover of thyroxine (thyroxine disposal or degradation rate) can be calculated from these observations. As shown in Fig. 5, the amount of thyroxine degraded per day falls from about 88 μg thyroxine iodine to about 42 μg between the ages of 20 and 90 years. Since the disposal rate is, at least in the steady state, equal to thyroxine synthesis these data provide evidence that secretion of thyroxine falls with advancing age (Gregerman, 1967).

There is, however, no evidence that the thyroid gland of aged subjects is unable to produce or release additional thyroxine when an adequate physiological stimulus is provided (Baker, Gaffney, Shock and Landowne, 1959). In these experiments 25 mg (10 USP units) of TSH were administered daily to two middle aged (46 and 51 years old) and three elderly (ages 81, 88 and 92) males for a period of 5 days.

The increments in pulse rate, basal oxygen consumption, protein bound iodine or I^{131} uptake by the thyroid gland were essentially the same for the middle aged and elderly subjects.

In summary, it may be said that there is no evidence that the thyroid gland shows any reduction in its capability to produce or release thyroxine as age advances. On the other hand, the daily degradation of thyroxine falls with age. However, animal experiments have failed to provide any evidence for an agewise reduction in oxygen uptake of tissue slices, homogenates or isolated mitochondria from rat heart, liver or kidney (Barrows, 1966).

There is no evidence that the amount of thyroxine required to regulate tissue and cellular metabolism is deficient in the aged. It should be noted that thyroxine degradation occurs largely in the liver and is not related to its effect in regulating cellular metabolism (Oppenheimer and Surks, 1971).

Since neither tissue metabolism nor the ability of the thyroid to produce thyroxine diminishes with age, what then is the basis for the progressive fall in total metabolism or energy production per M^2 surface area with age? In my opinion the fall is simply a reflection of the loss of metabolizing tissue with age.

As shown in Fig. 6 both height and weight diminish with age. The observations shown were made on 824 community residing males who are participants in a longitudinal study of aging (Stone and Norris, 1966). Participants in the study, who range in age from 20–96 years, spend 2 1/2 days in the hospital every 18 months for an extensive series of clinical, physiological,

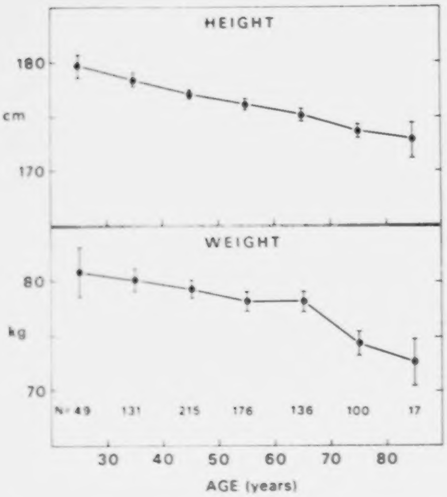


Fig. 6. Regression of height and weight on age in normal males.

biochemical and psychological tests. Since 470 of these subjects were measured three times or more between 1959 and June 30, 1969, it is possible to examine age trends within individual subjects of different ages. This was done by calculating the linear regression of height and weight on age for each subject. Mean values for the slope of the regression were then calculated for subjects within each decade. In Fig. 7, the average slope is plotted through the mean value for each age decade. It may be seen that, on the average, individual subjects also showed a decrement with age. In the case of body weight, subjects 50 years of age or younger showed, on the average, a gain in weight, but subjects 55 years and older showed an average loss in body weight. This longitudinal approach offers clear evidence of a gradual loss of body weight in subjects over the age of 55 years. The mean slope of the individual regressions of body weight on age is +0.236 kg per year in 45–54-year-olds, −0.146 for 55–64-year-olds, −0.487 for 65–74-year-olds and −0.385 for 75–84-year-olds.

Surface area is the traditional factor used to correct physiological measurements for differences in body size. In using this factor to reduce the effects of body size on measurements of basal energy production it is assumed that it is an effective index of the amount of metabolizing tissue in the individual. Since it is calculated from height and weight, it can be influenced by the amount of bone, fat and other tissues with low oxygen uptage which contribute to body weight, but not to oxygen uptake or energy production. Since the water content of cells is very stable and does not change significantly with age, it seemed reasonable to suppose that the water content of the body would serve as a more appropriate index of the amount of metabolizing tissue in the individual than surface area. Consequently, measurements of total body water were made in subjects by the antipyrine dilution

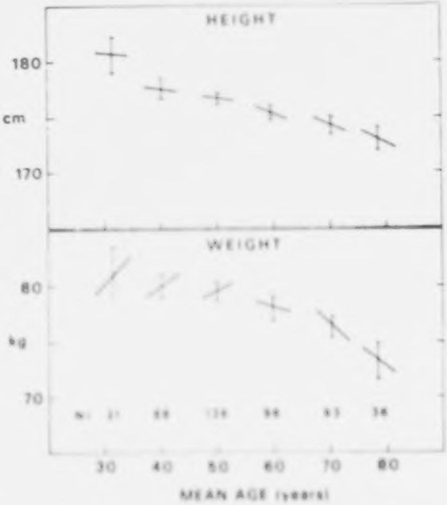


Fig. 7. Mean regression slopes of height and weight on age determined from serial measurements on the same subjects (normal males) over a period of 8 years. The vertical line represents ±1 standard deviation of the mean value.

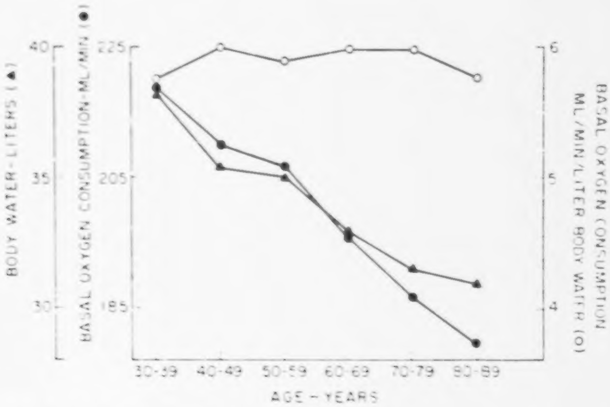


Fig. 8. Effect of age on basal metabolism in normal males. ▲, Total body water determined as antipyrine space (1); ●, Basal O_2 consumption, ml/min; ○, Basal O_2 consumption, ml/min/l body water. (Gregerman, 1967.)

technique (Shock, Watkin, Yiengst, Norris, Gaffney, Gregerman and Falzone, 1963). Fig. 8 shows that basal oxygen consumption per liter of body water is not influenced by age. It may, therefore, be concluded that the fall in basal metabolism (calculated as $cal/M^2/hr$) with age is simply a reflection of the loss of metabolizing tissue and all observations are compatible with the assumption that oxygen uptake of functional cells is not significantly diminished with advancing age.

Since the total energy production per 24 hours is the sum of the basal energy production and that required for daily activities, an attempt was made to determine age differences in daily activities and to relate total energy production to dietary intake. A total of 252 men, aged 20–29, participants in the longitudinal study, served as subjects for this study (McGandy, Barrows, Spanias, Meredith, Stone and Norris, 1966).

After detailed instructions from trained nutritionists each subject maintained a 7-day record of all foods eaten and mailed it to the Gerontology Research Center. After verification with the nutritionist, the data

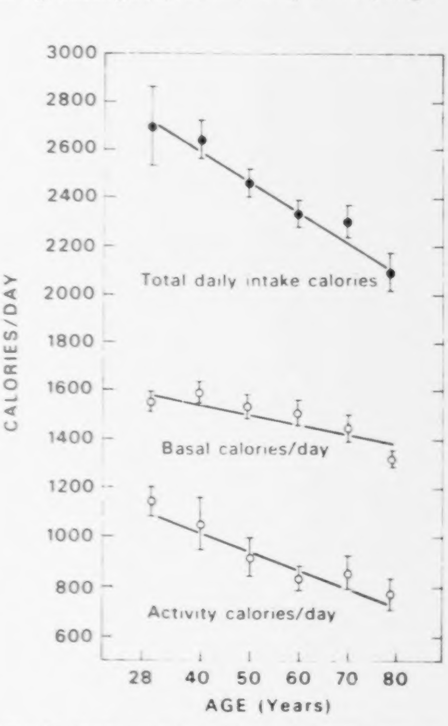


Fig. 9. Daily caloric intake and expenditure in normal males. ●—●, Total daily caloric intake; ○—○ (upper curve), Basal caloric expenditure per day; ○—○ (lower curve), Daily caloric expenditure for activity. Vertical lines represent ±1 standard curve of the mean. (Data from McGandy, Barrows, Spanias, Meredith Stone and Norris, 1966).

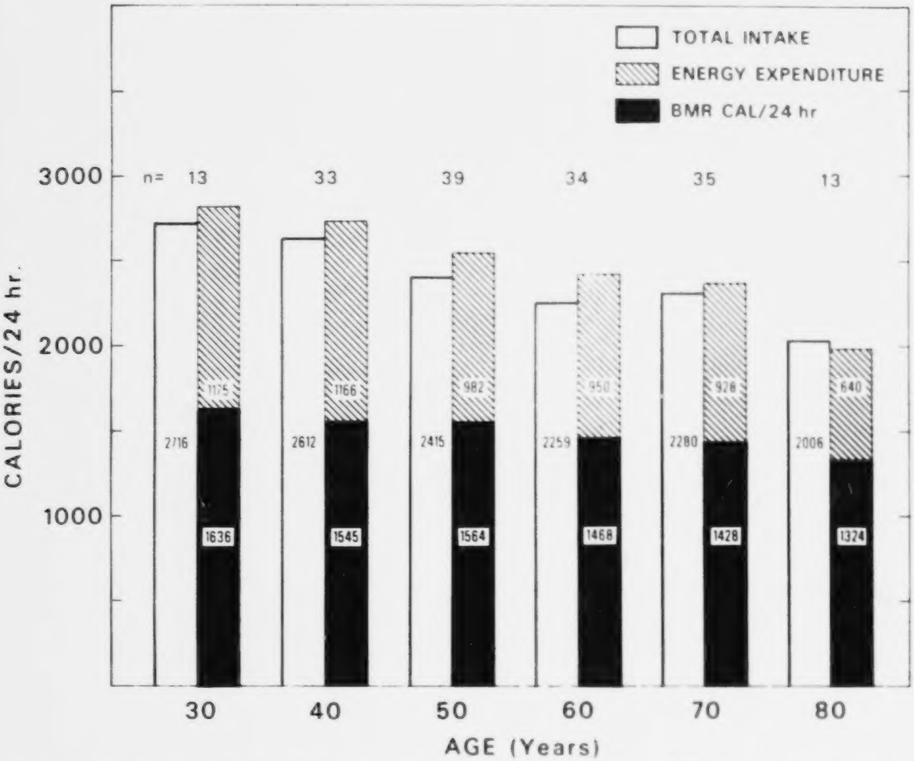


Fig. 10. Average daily energy balance in normal males aged 30–80 years. ☐, Total dietary intake cals/24 hrs; ■, Basal metabolism cals/24 hrs; ▨, Caloric expenditure per 24 hrs calculated from activity history. (McGandy, Barrows, Spanias, Meredith, Stone and Norris, 1966.)

from each record were coded and entered on punch cards for computer analyses of mean daily nutrient intakes.

Physical activity was estimated from a detailed interview covering specific activities at home, at work, at recreation (including active or passive participation in sports) and variations in activity patterns such as trips and seasonal sports. The amount of time spent in each activity was expressed as a daily average for each subject. Time spent in seasonal activities and activities which were pursued infrequently were expressed as an annual total and then divided by 365. Total daily energy expenditures were calculated for each subject by using values for each activity as reported in the literature.

As shown in Fig. 9 (top line), the total caloric intake in these subjects fell from 2 688 cal/day in 20–34-year-olds to 2 093 cal/day in 75–99-year-olds. The middle line of Fig. 9 shows the average basal metabolism expressed as cal per day for these subjects. The bottom line shows the total cals/day assignable to daily activities.

Of the 252 subjects who provided dietary records, 167 also submitted detailed records of their physical activity. It was found that this subsample did not differ significantly from the other subjects with respect to caloric intake or basal metabolism.

Fig. 10 shows the relationship between total caloric intake and energy expenditure (basal plus energy expenditure calculated from activity records) for males aged 30–80 years. A reduction in both caloric intake and total caloric expenditure with advancing age is shown. There is also fairly close agreement between caloric intake and expenditure at all ages. However, the calories required for activities fall more than basal calories, especially among the 80-year-old subjects. It is, therefore, clear that the reduction in energy metabolism in older subjects is a reflection of tissue loss and reduced activity.

SUMMARY

Aging is a phenomenon which proceeds over the entire adult life span. Its ultimate expression is in terms of a rise in mortality rate with age. Physiologically, aging is associated with a gradual decline in many physiological functions, such as cardiac performance, renal function, pulmonary function, etc. However, marked individual differences occur in the effects of age. The greatest age differences are found in performances which require the coordinated activity of a number of organ systems, as for example, physical exercise. Reserve capacities are reduced in the elderly and they require more time to adjust to displacing stimuli than do the young. The loss of reserve capacity is due in part to the loss of functioning elements in organs and tissues.

The reduction in energy metabolism in the elderly is due to loss of tissue and reduced activity.

REFERENCES

Andrew, W., Shock, N. W., Barrows, C. H., Jr and Yiengst, M. J. (1959). *J. Geront.*, *14*, (4), 405.

Arataki, M. (1926). *Amer. J. Anat.*, *36*, 399.

Baker, S. P., Gaffney, G. W., Shock, N. W. and Landowne, M. (1959). *J. Geront.*, *14*, (1), 37.

Barrows, C. H. (1966). In *Perspectives in experimental gerontology* (ed. N. W. Shock), pp. 169–181. Charles C. Thomas, Springfield, Ill.

Boothby, W. M., Berkson, J. and Dunn, H. L. (1936). *Amer. J. Physiol.*, *116*, 468.

Brandfonbrener, M., Landowne, M. and Shock, N. W. (1955). *Circulation*, *12*, (4), 557.

Buetow, D. E. (1971). Chapter 4 in *Cellular and molecular renewal in the mammalian body* (ed. I. L. Cameron and J. D. Thrasher), pp. 87–106. Academic Press, New York.

Davies, D. F. and Shock, N. W. (1950). *J. Clin. Invest.*, *29*, 496.

Gaffney, G. W., Gregerman, R. I., Yiengst, M. J. and Shock, N. W. (1960). *J. Geront.*, *15*, 234.

Gregerman, R. I. (1967). Chapter 8. In *Endocrines and aging* (ed. L. Gitman), pp. 161–173. Charles C. Thomas, Springfield, Ill.

Gregerman, R. I., Gaffney, G. W. and Shock, N. W. (1962). *J. Clin. Invest.*, *41*, (11), 2065.

Hügin, F., Norris, A. H. and Shock, N. W. (1960). *J. Geront.*, *15*, (4), 388.

Lewis, W. H., Jr (1938). *Amer. J. Physiol.*, *121*, 502.

McGandy, R. B., Barrows, C. H., Jr, Spanias, A., Meredith, A., Stone, J. L. and Norris, A. H. (1966). *J. Geront.*, *21*, (4), 581.

Miles, W. R. (1950). In *Methods in medical research* (ed. R. W. Gerard), Vol. 3, pp. 154–156. Chicago Yearbook Publishers.

Norris, A. H. and Shock, N. W. (1971). In *Science and medicine of exercise and sports*, 2nd edition. (In press.)

Norris, A. H., Shock, N. W. and Wagman, I. H. (1953). *J. Appl. Physiol.*, *5*, 589.

Norris, A. H., Shock, N. W. and Yiengst, M. J. (1953). *Circulation*, *8*, 521.

Norris, A. H., Shock, N. W. and Yiengst, M. J. (1955). *J. Geront.*, *10*, 145.

Oppenheimer, J. H. and Surks, M. I. (1971). Chapter 5 In *The thyroid* (ed. S. C. Werner and S. H. Ingbar), 3rd edition, pp. 52–65. Harper & Row, New York.

Rexed, B. (1944). *Acta Psychol.* Kbh, (Suppl. 33), pp. 1–206.

Shock, N. W. (1955). *J. Chronic Dis.*, *2*, (6), 687.

Shock, N. W. (1962). In *Proceedings of seminars 1959–1961* (ed. F. C. Jeffers), pp. 123–140. Council on Geront., Duke Univ., Durham, N.C.

Shock, N. W. and Yiengst, M. J. (1950). *J. Geront.*, *5*, 1.

Shock, N. W., Watkin, D. M., Yiengst, M. J., Norris, A. H., Gaffney, G. W., Gregerman, R. I. and Falzone, J. A. (1963). *J. Geront.*, *18*, (1), 1.

Silverstone, F. A., Brandfonbrener, M., Shock, N. W. and Yiengst, M. J. (1957). *J. Clin. Invest.*, *36* (3), 504.

Smith, L. E. and Shock, N. W. (1949). *J. Geront.*, *4*, 27.

Stone, J. L. and Norris, A. H. (1966). *J. Geront.*, *21*, (4), 575.

Suci, G. J., Davidoff, M. D. and Surwillo, W. W. (1960). *J. Exper. Psychol.*, *60*, (4), 242.

U.S. Department of Health, Education and Welfare, (1971). National Center for Health Statistics, Vital Statistics of the United States, 1968, *Vol. 2, Sect. 5, Life Tables, 1968.*

Yiengst, M. J. and Shock, N. W. (1962). *J. Appl. Physiol.*, *17*, (2), 195.

DISCUSSION

Carlson: Thank you, Dr. Shock, for a most important contribution to this important field. I would like to open the discussion after Dr. Shock's paper and there is one discussion in advance. Dr. Skinner, would you please make your comment.

Skinner: I would like to make a few comments regarding physical activity and aging, as seen from another point of view than that of Dr. Shock. In a study on a total population in Tecumseh, Michigan (about 9 000 persons), the leisure activity of a subsample of 1 965 persons was estimated from a very detailed physical activity questionnaire by Cunningham et al. In Figure 1 one can see the number of hours per week which the subjects participate in leisure activities relative to age. It can be clearly seen that there is a reduction in the hours of activity per week among the older subjects. Of more importance, however, is the

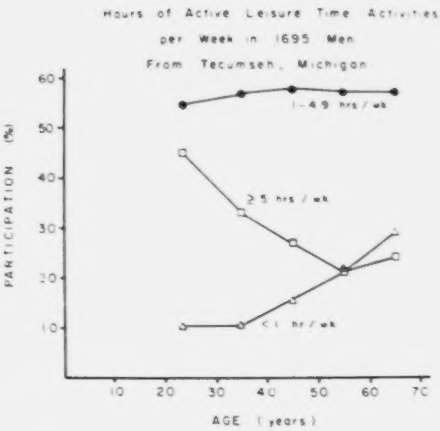


Fig. 1

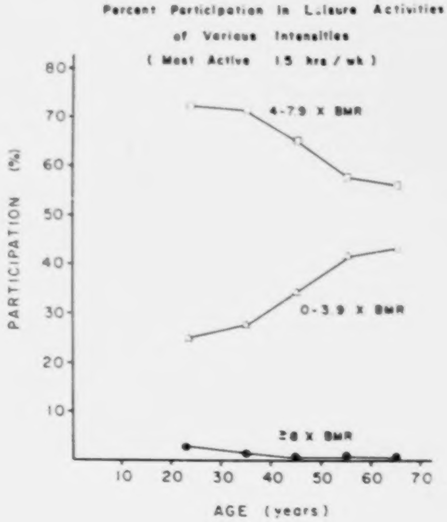


Fig. 2

Fig. 1 and 2. Modified from data of Cunningham, et al., 1909. Cunningham, D. A., H. J. Montoye, H. L. Metzer and J. B. Keller. Physical activity at work and leisure as related to occupation. *Med. Sci. Sports* 1: 165–170, 1969.

intensity of the activities in which they participate (Fig. 2). Looking at the intensity of the "most active" 1.5 hours per week, there is a marked shift in the pattern of participation so that in the older age range about 50% of the subjects are doing activities requiring only 1–3.9 times the basal metabolic rate (sleeping to walking). Thus, the older persons are not only participating less often but the intensity of their activities is also lower.

The intensity of an activitiy can also be viewed in another way. Rather than classifying intensity in an absolute sense (kcal or multiples of the BMR), it is also possible to do it in a relative sense, i.e. the intensity of a given activity is relative to the maximal intensity which a person is able to perform. Since the maximal working capacity (endurance, strength, etc.) decreases with increasing age, while the energy cost and heart rate response to standardized submaximal work loads remains essentially the same, for the same caloric expenditure the older person is working at a higher percentage of his maximal capacity.

Another point I would like to mention comes from the psychological studies of older persons by Kreitler and Kreitler (1). Older persons tend to have a poor self-image relative to their ability to move. They then reduce the amount of activity they do, thus reinforcing their poor self-image. These investigators found that after this cycle was reversed by putting these older persons into an exercise program, the self-image was improved and the subjects tended to be more active.

Carlson: Dr. Shock, you have described two features of aging i.e. loss of cell mass and decline in physical activity. Do you know which of these that comes first? Or to re-phrase it, is decline in cell mass causing elderly people to be physically less active or is it because people with increasing age are becoming more lazy that the cell mass declines with age?

Shock: We have no direct evidence as to which comes first. However, the age changes in chemical composition that appear in rat muscle are similar to those that appear in young animals in which the muscle is immobilized or denervated. It is therefore possible that the age changes in muscle may be due in part to an atrophy of disuse. However, it is impossible that the loss of nephrons from the old kidney can be attributed to atrophy of disuse.

Munro: I want to ask two questions, one of Dr. Shock and one of Dr. Skinner on the basis of what he said. The question to Dr. Shock is on cell loss. Does cell loss in old age mainly occur in tissues no longer showing active cell division, or in dividing tissues?

For Dr. Skinner I would like to ask: In old age, does efficiency of work output per calorie change?

Shock: The cell loss to which I referred occurs only in non-dividing cells. In epithelial cells, division continues in elderly subjects, but the rate of cell division may be sligthly reduced in old age.

In our normal subjects we have found no age differences in efficiency at moderate work levels. At very slow rates of work, efficiency is somewhat lower in old than in young subjects. At maximum work rates, efficiency is also lower in the old than in the young—but the old subjects cannot obtain as high a rate of work as do the young.

Skinner: If I remeber correctly, Dr. Munro, Taylor and his associates in Minneapolis found a slight (approximately 5%) reduction in efficiency of walking on a treadmill among older workers of the railroad industry. Other evidence seems to suggest that there is little or no change in efficiency with age.

Gopalan: Dr. Shock has concluded that the reduction in basal oxygen consumption in old age is totally attributable to reduction in the metabolising tissue, and that no reduction in oxygen consumption per unit of metabolising tissue is involved. This conclusion is based on his finding that the reduction in total body water and the reduction in basal oxygen consumption run almost parallel. If the relationship between total body water and metabolising tissue undergoes no change during ageing, this conclusion may be testified. On the other hand, if the total body water in relation to metabolising tissue tends to decrease with age, then computations based on total body water may tend to underestimate metabolising tissue and thus mask the reduction in oxygen consumption per unit of metabolising tissue. In chronic starvation there is reduction in basal oxygen consumption. Our studies on cases of starvation in India as well as the studies of Keys and colleagues on induced starvation in human volunteers go to show that this reduction is to a great extent attributable to reduction in metabolising tissue and to a small extent to reduction in oxygen consumption per unit of metabolising tissue.

Shock: The original studies of Lawrey and Hasting as well as our own failed to find any evidence of "tissue dehydration" in muscle tissue from old rats. There is, however, a reduction in the amount of intracellular water and an increase in extra-cellular water in skeletal muscle tissue in old rats. However, the intracellular water, per kg of cells, remains remarkably constant with age. Hence we originally used intracellular water as the index, *that* is the difference between anti-pyrene and eterocyanate space. However, because of the variance introduced by calculating the difference between two measurements, we have found that using total body water reduces the variance of the index. Since isolated malochondria from cardiac muscle or liver from rats show no age differences in oxygen uptake, it seems reasonable to assume that no age differences in total metabolism of cells exist. Since comparable data are not available on human tissues, we can only assume that similar results would be found.

Nordquist: You mentioned that nerve conduction velocity decreases with age. This must be solely attributed to changes in the nerve potential of the cell membrane at the Rauvier's nodes of the A-fibres, if you use the common method measuring conduction velocity of n. ulnaris. My question is: does the action potential spike have a different shape when recorded from young and old individuals. If it is broader in old people it means to me, that the different fibres, which compose the action potential spike of the nerve are in a more variable state of metabolism in old people than in young people.

Shock: The method we used for determining nerve conduction velocity in human volunteers did not give action potential spikes of single nerve fibres which would be necessary for this type of analysis and interpretation.

Vahlquist: Dr. Shock, one of your slides showed that with decreasing age there is a decrease not only in weight but also in height. Do you feel that the latter observation to any extent could be influenced by the "secular trend"?

Shock: At present we do not have a good explanation for the decrease in height. From serial observations made on the same subject over a 10 year time interval we can say that the decrement is a real one and is not due to sampling differences. It is no doubt true that our 70-year-old subjects were shorter at age 18 than are 18 year olds at the present time, but I doubt whether this will explain the reduction in height which we have observed.

REFERENCE

1. Kreitler, H. and Kreitler, S.: *In* Physical Activity and Aging (ed. D. Brunner and E. Jokl), Karger, Basel, 1970.

Orimo, H.; Shimada, K.; Iriki, M.; and Maeda, D. (Eds.),
Recent Advances in Gerontology. XI International Congress of Gerontology
Amsterdam: Excerpta Medica, 1979

Patterns of longitudinal changes in renal function

N.W. SHOCK, R. ANDRES, A.H. NORRIS AND J.D. TOBIN

Gerontology Research Center, National Institute on Aging, Baltimore City Hospitals, Baltimore, MD, U.S.A.

The goal of this symposium was to demonstrate that the analysis of observations collected longitudinally brings to light information about aging that could not be derived from cross-sectional data alone.

The Baltimore Longitudinal Study of Aging, initiated in 1958, is based on a population of approximately 650 community residing males who spend 2.5 days at the Gerontology Research Center every 1.5–2 years for an extensive series of biomedical and psychological tests. Cross-sectional analysis of measurements of standard creatinine clearance in 548 normal males, aged 25–100 years, showed a progressive linear decline from 140 ml/min/1.73 m^2 at age 30 to 97.0 ml/min/m^2 at age 80 (Rowe et al., 1976). The present report deals with the results of 24 hour creatinine clearance tests on 398 subjects aged 25–100 years who were tested five or more times over a period of 10.0 ± 0.15 (SE) years.

Linear regressions of creatinine clearance on age were calculated for each subject. The regression slopes ranged from +8.0 to −9.2 ml/1.73 m^2/min per year, mean −.63 ± .109. The correlation of the individual regression slopes with age was highly significant ($r = -.171$, $P < 0.001$). Average slopes for 20-year age groups decreased progressively from −.26 ± .331 for the 20–39 age group to −1.51 ± .504 for subjects aged 80–100 years (Table 1). This analysis of longitudinal data shows that the rate of fall in creatinine clearance increases with age and is not linear over the age range 25–100 as would be inferred from cross-sectional data.

Although the standard error of individual regression coefficients based on as few as five measurements may be large, eight subjects were identified with statistically significant positive coefficients (b ranged from 2.0 to 6.3). Longitudinal observa-

Table 1 Mean values of individual regression slopes by age

Age (yr)	N	Mean slope ($\bar{b}$ ± SE) (ml/min/1.73 m^2/yr)
20–39	33	−.26 ± .331
40–59	201	−.40 ± .145
60–79	149	−.92 ± .193
80–100	15	−1.51 ± .504
20–100	398	−.627 ± .109

Table 2 Age distribution of subjects with slopes > +1.0

Age (yr)	N		Percent
	Total	b > +1.0	
25–34	11	2	18.2
35–44	64	13	20.2
45–54	110	26	23.6
55–64	88	11	12.5
65–74	80	7	8.8
75–84	42	3	7.1
85–100	5	0	0

tions can identify individual subjects who deviate markedly from the average pattern of age changes derived from cross-sectional data.

Table 2 shows the distribution by age decades of individuals with regression slopes greater than +1.0. After age 45 the percentage of subjects showing positive regression slopes greater than +1.0 fell from 23.6% to 0 at ages 85–100 years.

Five or more observations at 1 to 2 year intervals were available on 59 subjects who died after the age of 55. Death certificates showed that 2/3 of the subjects died from cardiovascular diseases and 1/3 from cancer and other causes. Renal disease was not recorded as the cause of death in any of the subjects. Mean values for creatinine clearances and individual regression slopes for this group of 59 subjects were compared with values based on a group of 106 subjects over the age of 55 years who are still living. Table 3 shows that the mean age of the two groups did not differ significantly. However, the mean clearance level and the average of individual regression slopes of creatinine clearance were significantly different. The subjects who died had a lower creatinine clearance than the living subjects of the same age and the rate of fall in clearance was significantly greater in those who died than in those who remained alive.

In summary, these studies show that longitudinal data permit (1) identification of the true characteristics of age *changes* in individuals (cross-sectional studies can only provide information on age *differences*), and (2) identification of individual patterns of age changes. Even though there is, on the average, a decrement in clearance decade by decade even in early adult life, the longitudinal technique

Table 3 Comparison of living and dead subjects (subjects aged 55—100 years)

	N	Age (yr)	Mean level (ml/min/1.73 m^2)	Slope (b ± SE) (ml/min/1.73 m^2/yr)
Dead	59	72.0 ± 1.04	100.3 ± 3.26	−1.58 ± 0.327
Alive	106	70.1 ± 0.67	108.2 ± 1.64	−0.70 ± 0.206
Difference		1.9 ± 1.24	−7.9 ± 3.65	−0.88 ± 0.386
t		1.53	2.165	2.277
P <		NS	0.05	0.025

shows that there are some individuals who show a remarkable maintenance of their renal function. Some individuals show improvement in renal function with age. Subjects who died after the age of 55 had lower clearance values and showed a greater rate of decline in renal function over the ten years preceding death than did those who are still living.

REFERENCE

Rowe, J.W., Andres, R., Tobin, J.D., Norris, A.H. and Shock, N.W. (1976): *J. Geront.*, *31*, 155.

Danon, D., Shock, N.W., and Marois, M. (Eds.),
Aging: A Challenge to Science and Society, Vol. 1, Biology
New York: Oxford U.P., 1981

Physiological Indices of Aging

Jordan D. Tobin

The concept of a physiological index of age is intuitively appealing since we have often said that someone 'looks younger than his age' or 'performs better than people his age'. Inherent in these statements is the implicit notion that we know how someone of a given age should look or perform, that we know what is normative or standard. Within the discipline of human physiology, this determination is frequently difficult. In addition to the problems of subject selection, applicability of a given sample to another population, standardization of test conditions, and the choice of which physiological system to study, there is the almost philosophical question of the interrelationships of age, physiology, and disease.

Previous studies, as recently reviewed by Shock (1978) and Costa (1977) have attempted to develop a physiological or functional age for individuals using a multiple regression model. The variables used have included physiological, anthropometric, physical, and biochemical indices. Performances on these tests were used to predict a 'functional' age.

A different approach has been taken for this study. Four physiological variables which are considered clinically important in that they measure functions which are related to the health of an individual were examined. These included the respiratory system (forced expiratory volume in 1.0 second), the renal system (standard creatinine clearance), the cardiovascular system (systolic blood-pressure), and metabolism (oral glucose tolerance). Each of these variables not only is influenced by age but in addition is associated with a disease (chronic obstructive pulmonary disease, renal failure, hypertension, and diabetes). Normative data were derived from individuals who were free of diseases or medications known to influence the system being studied. Age-adjusted standard T-scores were then calculated for all individuals. In order to judge the importance of each variable, the age-adjusted T-scores for these individuals who have died were compared to those who lived.

Population

The data on this study were derived from tests on the volunteers of the Baltimore Longitudinal Study of Aging conducted within the National Institute on Aging under the direction of Dr. Nathan Shock and Dr. Reuben Andres. The study had its inception in 1958 when a retired Public Health Service physician came to Dr. Shock and suggested that researchers on aging of humans should not be studying only the residents of nursing homes and chronic disease hospitals, who represent a minority of the aged population, but should also be looking at the 'healthy' aged. Towards this aim, he offered himself for study and agreed to recruit his friends as well. Thus was formed the nucleus of the study population which has continued to be a self-recruited group of males ranging in age from 18 to 103, with over 1100 men having been seen once, and a currently active group of 650 community-dwelling volunteers.

The characteristics of the group have been previously described (Stone and Norris 1966): briefly, they are predominantly highly educated, upper middle class, white, Protestant, and in academic, managerial, or government positions. No subject is excluded from the study for health reasons. They are admitted for 2½ days to the Baltimore City Hospitals and undergo a series of more than 40 medical, physiological, and psychological tests. They have agreed to return to be retested at 2-year intervals until they are 60 years old, at 18-month intervals until they are 70 years old, and then to return yearly for the remainder of their lives. These dedicated volunteers form the data base for the results to be presented and, obviously, we owe them a great deal of thanks.

We have mentioned that no subject is excluded from the study on the basis of health. However, when we are analysing the effect of *age* on a physiological variable, for instance, the glucose tolerance test, we would not want to include in the group of normal individuals those with the disease diabetes mellitus. These diabetics would have poor performance on this test of metabolism, which would not be a function of their age but of their disease. Similarly, we would not include anyone taking drugs known to influence carbohydrate metabolism or having other diseases which influence it. Thus, for each variable studied a clinical 'clean-up' is necessary in order to ascertain the effect of age *per se*, rather than disease.

Glucose metabolism

The oral glucose tolerance test is a physiological test of metabolism. A blood sample for analysis of glucose is taken after an overnight fast under basal conditions. The subject then drinks a solution of glucose (1.75 grams per kg of body weight) which represents a challenge to the metabolic system and, in part, simulates what each of us does during the day as we eat. Performance on this test is judged by how efficiently the subject metabolizes the load of glucose and returns his blood glucose concentration towards the fasting level. Towards this aim, the blood glucose concentration at two hours is clinically used to categorize people as 'normal', 'borderline', or 'diabetic'.

Results in a 'clinically clean' group of subjects indicate that there is no effect of age on the fasting (unstressed) glucose level. After the stress of the glucose load, initially there is no age effect discernible as the glucose is absorbed from the gut and the blood glucose rises equally in all age groups for the first 40 minutes. By one hour, the 20-year-old subjects have already started to lower their glucose level, and by two hours, there is a clear age-ordering of concentration. The 20-year-olds have the lowest glucose, the 30-year-olds next, the 40-year-olds have not performed as well as the 30-year-olds, but they are better than the 50-year olds, etc with the 80-year-olds having the highest glucose level and, therefore, the poorest performance. The two-hour plasma glucose concentrations for each decade are shown in Fig. 1. These results are mean values; there is a large variance at any age, and there are clearly some superior 70-year-old subjects who behave as well or better than the *average* 20-year-olds. Since these are cross-sectional results, we do not know if these super-performers were even better performers when they were younger and have, in fact, deteriorated as they aged or if they have maintained the same level of performance throughout their life span.

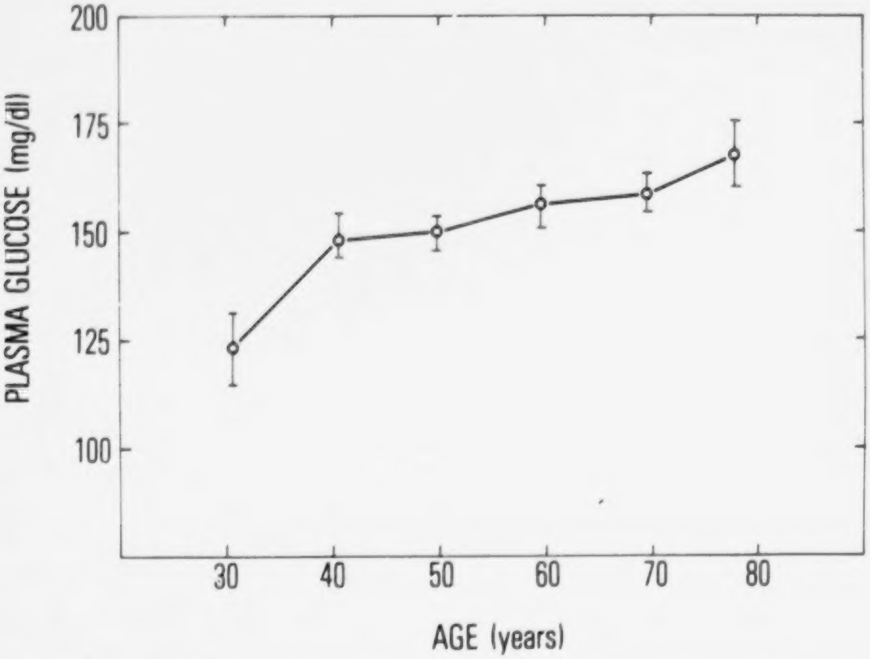


FIG. 1. Effect of age on plasma glucose concentration 2 hours after oral glucose. Clean group. The *N*s for each decade from 30–80 were 14, 50, 82, 67, 53, and 33.

Were the usually accepted criteria of performance on this test (derived from young people) applied to this health group of subjects, more than 50 per cent of the men over the age of 60 would be classified as 'diabetic'. It is worth emphasizing once again that these subjects represent a 'clean' group. They have no family history of diabetes, no diseases known to influence carbohydrate metabolism, and are taking no medications that would influence their performance. They are active, healthy, and, by dietary diary, taking adequate amounts of carbohydrate in their diet. The finding of such a high prevalence of 'abnormal' results on glucose tolerance testing is not in keeping with the known prevalence of the disease, diabetes.

We have chosen (Andres 1971) to judge performance on this test using an age-adjusted nomogram constructed from these data. With this technique, a subject can be judged against his age peers, and a percentile rank can be assigned to his performance. An exactly average performance at any age will have a 50 per cent rank; a 5 per cent rank indicates that only 5 per cent of subjects of that age perform that poorly. The actual glucose level that determines a rank of, for example 5 per cent, is of course higher in the 70-year-olds than in the 20-year-olds. The nomogram does not indicate what is normal and what is abnormal but does allow more flexible and appropriate (but equally arbitrary) judgements than one arbitrary diagnostic cut-off level for all ages.

In order to determine which percentile ranking at different ages is significant in terms of predicting future health problems, prospective longitudinal studies are required. A variety of 'end-points' need to be examined, end-points known to be associated with diabetes mellitus. Thus, since mortality rates are markedly increased in diabetics, various levels of glucose tolerance performance should be examined for correlation with mortality. Similarly, the development of the known complication of diabetes, such as coronary heart disease and the micro-angiopathies (eye, peripheral nerve, and kidney problems) should be analysed. Finally, the development of florid diabetes is an essential end-point to be analysed.

Pulmonary function

The functions of the lung in terms of gas exchange, ridding the body of carbon dioxide, and supplying oxygen, show no age effects in the basal state. There are no differences in the content of oxygen, carbon dioxide, or electrolytes in the blood. There are marked age effects, however, on tests of the ability of the lung to perform when stressed. The amount of air that can be forcibly expelled in one second (Forced Expiratory Volume, FEV 1.0) has been found useful as a clinical test of respiratory performance. Results of these tests are profoundly influenced by diseases of the lung (bronchitis) and by smoking. In order to assess the effect of age on this test, therefore, another 'clinical clean-up' was necessary. Individuals who were free of overt pulmonary disease, were non-smokers, and who had no other diseases or who were taking no drugs which might affect performance, formed the group of healthy subjects in this analysis. They showed a progressive decrease in performance across the age span, with each successive age decade performing less well than the younger decades (Fig. 2). Again, it is not clear whether the poorer average performance of the older subjects on this test of pulmonary performance is a physiological age effect on all subjects or if it is indicative of the development in some of the subjects of subclinical disease.

Blood-pressure

The control of the blood-pressure level is a complex interrelationship of anatomical, hormonal, neural, renal, and cardiovascular factors. We are examining the summation of these factors when we measure this variable, and this must be kept in mind. The systolic blood-pressures reported here were obtained on our volunteers during their physical examinations. All subjects were screened and any with cardiovascular, renal, or metabolic diseases were excluded from this analysis, as were those on any medications (diuretics, antihypertensives, etc.) known to influence blood-pressure.

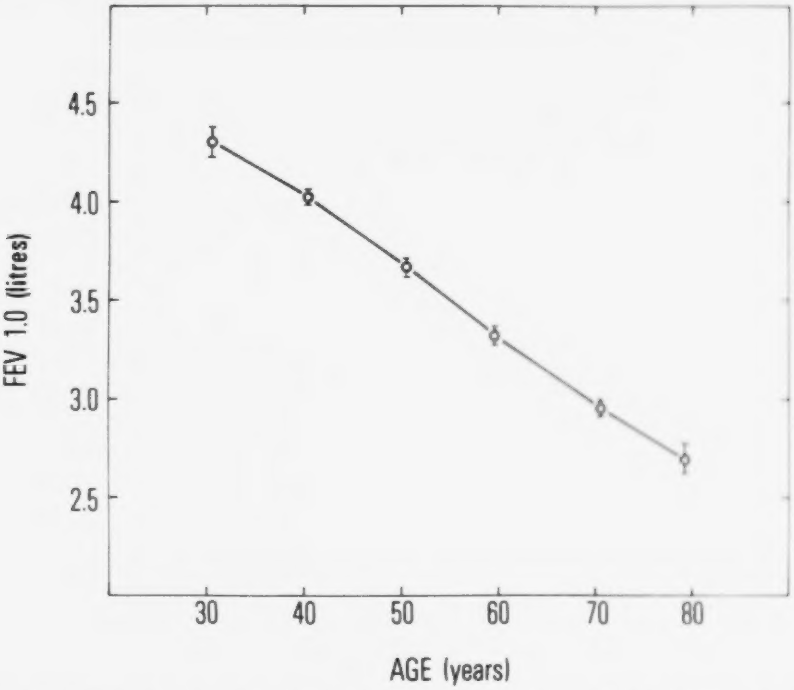


FIG. 2. Effect of age on forced expiratory volume at one second. Clean group. The *N*s for each decade from 30–80 were 97, 107, 151, 114, 115, and 45.

There was a significant increase of systolic blood-pressure with age, with each succeeding decade having a higher average pressure than the preceding one (Fig. 3).

Renal function

One measure of the ability of the kidney to function is the creatinine clearance, which is an estimate of the glomerular filtration rate. It measures how many

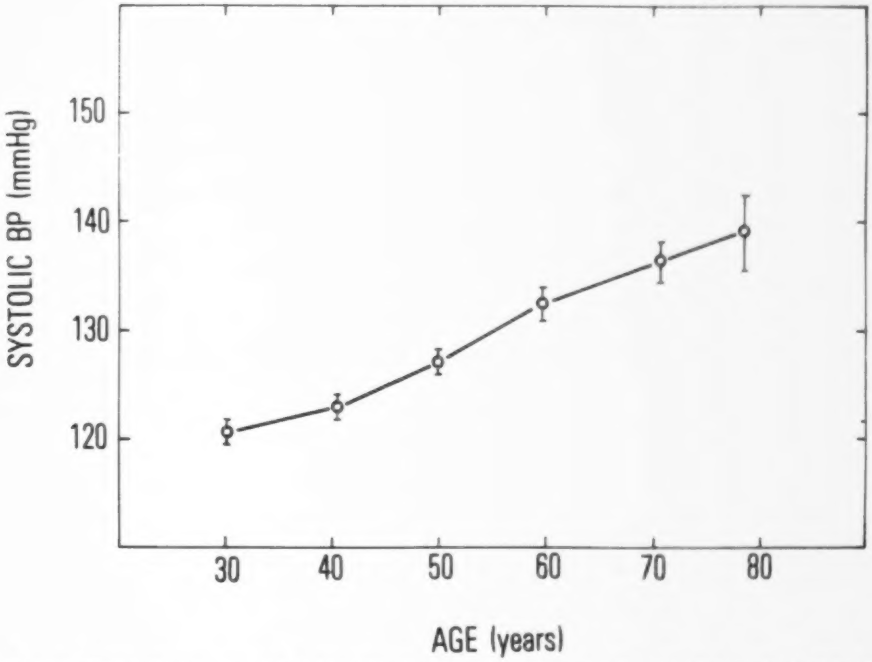


FIG. 3. Effect of age on systolic blood-pressure. Clean group. The *N*s for each decade from 30–80 were 82, 151, 184, 119, 103, and 35.

millilitres of plasma are 'cleared' of nitrogenous wastes each minute, with higher numbers (adjusted for body size) indicating better function and lower numbers poorer function. The results of a cross-sectional and longitudinal analysis of creatinine clearance have been presented (Rowe, Andres, Tobin, Norris, and Shock 1976). There is a highly significant decrease in clearance across the age span, with each succeeding decade being lower than the preceding one (Fig. 4). These results are on a clinically clean group who have no diseases known to influence renal function, are on no medications which would influence performance, have not had prostatectomies, and who have a normal urine analysis and no history of renal disease.

Analysis

In these four examples of physiological variables which are considered to be clinically important in medicine, there was a decrement in function with age. This was in a group of individuals who were thought to be free of significant disease within the limits of medical judgement; and the decline cannot simply be ascribed to 'sick old men'. The importance of this decline, however, remains a question. Should two individuals who have the same level of performance be considered to be equivalent in a functional sense, even though one of them (a young man) reached this low level because of a disease, while the other (an old man) appears free of disease and has reached this level because of a physiological decrement with age? One approach to this question is to use an age-adjusted score for the subject, and to test whether performance is related to the future health or survival of the individual. An age-adjusted T-score was calculated for each individual. The mean value for each decade was assigned a value of 50 with a standard deviation of 10. Thus, an individual who was one SD above the mean for his age group would have a T-score of 60 (50 + 10), while someone one and one-half SDs below the mean would have a T-score of 35 (50 − 15). For convenience, good performance was always expressed as a T-score above 50 (for BP and glucose tolerance, the better performers, in fact, have been lower absolute values and these were reversed). The T-score thus serves two functions: (1) it removes dimensions from consideration and expresses different variables in the same units, and (2) it age-adjusts performance since each individual is judged against the mean and standard deviation of the performance of those clinically 'clean' individuals in his own age decade.

Since the initiation of the Baltimore Longitudinal Study, 162 of the volunteers are known to have died. T-scores were calculated for each individual's performance on each variable and the scores for those individuals who subse-

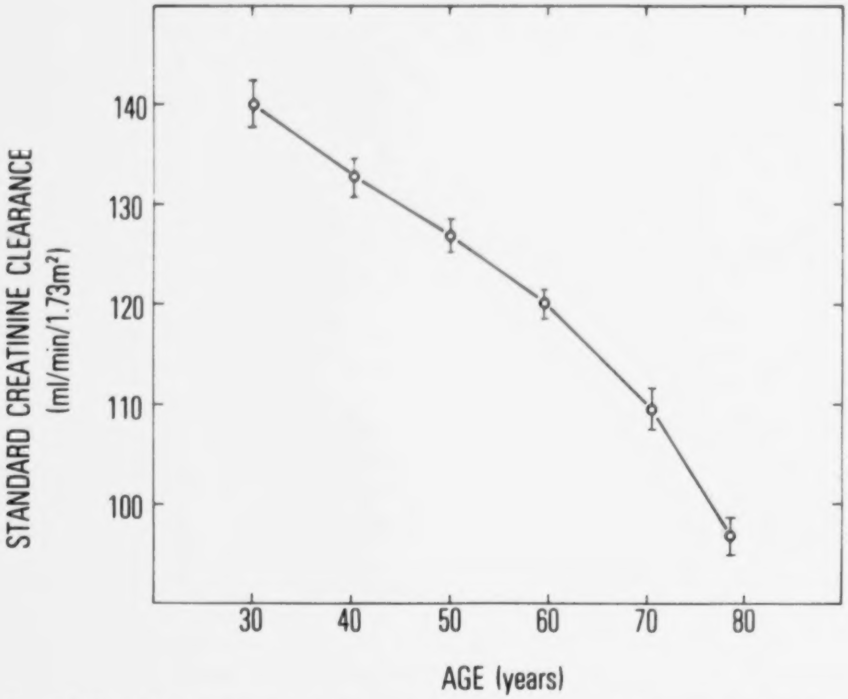


FIG. 4. Effect of age on creatinine clearance. Clean group. The *N*s for each decade from 30–80 were 73, 122, 152, 94, 68, and 29.

quently died were compared to those who lived. Fig. 5 graphically shows this comparison for each variable.

For the first three variables, blood-pressure, FEV 1.0, and creatinine clearance, the T-score of those volunteers who lived was not significantly different from a score of 50.0 (the mean score of the clean group). The T-scores of those who died, however, were all significantly lower, and represent a poorer age-adjusted performance on the average for those variables (Table 1). There was no significant difference between the group that lived and the group that died on the fourth variable, the glucose concentration at two hours of a glucose tolerance test. Both groups, however, had mean T-scores lower than 50. This difference (which was not observed in the other three variables) probably represents the greater sensitivity of the glucose tolerance test to those factors, especially medications, used to exclude subjects from the clean group.

Conclusion

Normative data were derived on men who were free of diseases which might influence the results. The decrements of physiological functioning of the four variables studied (representing four different systems) are not a result of diseases or medications, and must be considered a function of age. Thus, kidney function

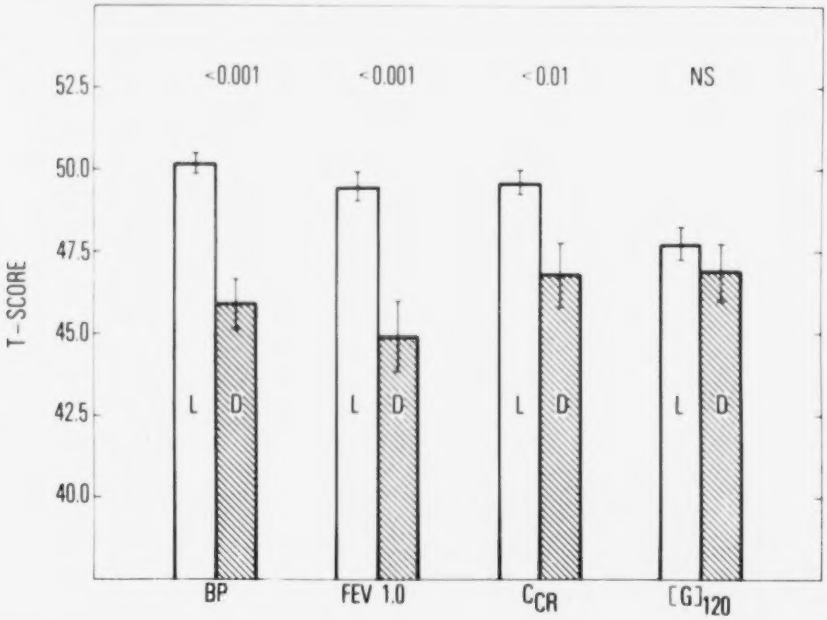


FIG. 5. Relation of performance level on four physiological tests to survival. Cross-hatched bars (D) are results on those subjects who have died; open bars (L) are results on the survivors. BP = systolic blood-pressure on physical examination. FEV 1.0 = forced expiratorycapacity in one second. C_{cr} = creatinine clearance. $[G]_{120}$ = glucose concentration 120 minutes after oral glucose. T-score (see text).

Table 1 T-scores for live and dead groups

		BP	$FEV_{1\cdot 0}$	C_{cr}	Glucose
Live					
	Mean	50.2	49.4	49.6	47.8
	SEM	0.33	0.37	0.35	0.46
	N	860	813	772	676
Dead					
	Mean	45.9	44.9	46.8	46.9
	SEM	0.77	1.08	1.01	1.20
	N	162	135	155	100
p		<0.001	<0.001	<0.01	NS

decreases with age, not because there is more clinical renal disease in the elderly but because of age and the passage of time. The actual mechanism of the decline of a system, be it loss of nephrons in the kidney or a decreased sensitivity of the pancreas to respond to glucose and secrete insulin in the decline of glucose tolerance with age, is under investigation. There may be a central, uniting aging factor with different expression in each organ system but, at the present time, this must be considered speculative.

Just as the mechanisms responsible for the demonstrated age changes are unclear and require future research, the resultant effects of the decline in a practical sense also need further study. In three of the four variables reported, those volunteers who have died had significantly poorer age-adjusted performance than those who lived. Total mortality is, however, a crude (though definitive) end-point. In this population, the bulk of the deaths were, as expected, due to cardiovascular and malignant disease. As the study progresses, more specific mortality data will be available and allow comparison of age changes in function with disease entities. The relationship of poor performance to the development of disease is not a simple one. Does the decreased function and the decreased reserve capacity to respond to stress serve as a fertile soil for disease to seed and grow? Do the organ systems progressively lose more and more function until some critical point is passed and they are, in effect, worn out? The list of questions goes on and, hopefully, the list of answers will grow. The answers will serve not only to increase the body of knowledge about the physiology of aging but, also, an understanding of the processes of health and function in the elderly.

References

ANDRES, R. A. (1971). *Med. Clin. N. Amer.*, **55**, 835.

COSTA, P. J. *Proceedings of the Second Conference on Epidemiology of Aging*, DHEW Publication. (In press.)

ROWE, J. W., ANDRES, R., TOBIN, J. D., NORRIS, A. H., and SHOCK, N. W. (1976). *J. Gerontol.*, **31**, 55.

SHOCK, N. W. (1980). This volume.

STONE, J. L. and NORRIS, A. H. (1966). *J. Gerontol.*, **21**, 575.

J Appl Physiol 45: 536–539, 1978

Longitudinal changes in basal metabolism in man

STEPHEN P. TZANKOFF AND ARTHUR H. NORRIS
Gerontology Research Center, National Institute on Aging, National Institutes of Health, Baltimore City Hospitals, Baltimore, Maryland 21224

TZANKOFF, STEPHEN P., AND ARTHUR H. NORRIS. *Longitudinal changes in basal metabolism in man*. J. Appl. Physiol.: Respirat. Environ. Exercise Physiol. 45(4): 536-539, 1978. – In a recent cross-sectional study of aging in adult men all of the age-related differences in whole-body basal Vo$_2$ (WBVo$_2$) were attributable to differences in skeletal muscle mass (creatinine excretion). This study sought to extend those findings on 355 adult men on whom five or more paired determinations of WBVo$_2$ and muscle mass were obtained over a mean of 10.7 yr. Individual rates of change were calculated by the least-squares method. The overall mean was -0.82 ml O$_2$ · min^{-1} · yr^{-1} and was similar to cross-sectional data. Mean slopes of WBVo$_2$, summarized by age-decade groups, were all negative, not different from one another, and consistent with the cross-sectional trend. However, slopes for nonmuscle Vo$_2$ (NMVo$_2$) were positive for the six older groups but not consistent with the unchanging cross-sectional trend. Reassessment of the subject population revealed that 48 men had died some time (mean 1.9 yr) after their last measurements. Cancer and cardiovascular disease accounted for 46 deaths. As a group all decedents had significantly higher mean slopes for NMVo$_2$. Muscle mass decrement and the resulting decrease in aerobic requirement accounted for aging decrements in WBVo$_2$ in the absence of these terminal diseases. When these conditions were present the overall decline was slowed or even reversed by gradual increases in NMVo$_2$. In longitudinal studies this may be mistakenly interpreted as a stabilization of the aging trend.

human aging; basal Vo$_2$; basal metabolic rate; creatinine excretion; muscle mass; body composition; gerontology

---

BODY SIZE, in units of surface area computed by the standard equation of DuBois and DuBois (2), is the traditional reference for the basal metabolic rate (BMR). Body size, however, ascribes equal weight to adipose, bone, and connective tissues, which have very low aerobic demands, and all other organs, which, although very different from one another in aerobic demands relative to their mass, account for the bulk of the whole-body basal O$_2$ consumption (WBVo$_2$).

At the turn of the century Rubner (9) suggested that the "active tissue mass" was a more suitable reference for the BMR. However, practical considerations severely limit the measurement of such mass. There has been general agreement that fat-free or lean body mass more closely quantifies the mass of active tissues (1, 4, 6); however, indirect measurements of this quantity are, at best, difficult, fraught with assumptions, and impractical for general use.

In a recent report on a cross-sectional analysis of aging (12) we showed that, in adult men, the age-related decrements in the rate of WBVo$_2$ were wholly attributable to concurrent decrements in the mass of metabolically active, creatinine-producing skeletal muscle. Mean Vo$_2$ of the noncreatinine-producing tissues, which for pratical purposes represent the aggregate of all metabolically active organs except skeletal muscle, did not differ in any statistically significant manner among all age groups studied.

Our cross-sectional analysis results suggest that aging in adult men is accompanied by gradual loss of skeletal muscle. Other metabolically active tissues and organs, as represented by the aggregate of the noncreatinine producing tissues, undergo little or no change in their basal O$_2$ requirements. However, a valid criticism of cross-sectional study designs, particularly for those extending into old age, is that they reflect selective mortality (5). Longitudinal studies, although expensive in time and cost, more accurately describe changes with age and avoid some of the problems inherent in the cross-sectional designs.

Except for scattered observations on individuals as they aged, the only systematic longitudinal analysis of aging on the basal O$_2$ requirements of men is a report by Keys et al. (5). They interpreted their data on body density (as in index of body fatness) and WBVo$_2$ as evidence that the measured decrement in basal O$_2$ requirement over the ages 20–75 yr was related to changes in body composition and very little to aging itself. This interpretation is consistent with our more specific cross-sectional findings (12) but lacks reliability in that the oldest group of men in which body density was measured averaged only 49.8 yr of age.

The present study examines the longitudinal changes in Vo$_2$ over the whole adult age range for each of the two oxygen-consuming compartments, muscle and nonmuscle, which were described in the earlier cross-sectional report. This approach is advantageous in that it need not account for either the deposition or loss of fat, nor the questionably valid reference to body size.

METHODS

Subjects for this study were participants in the Baltimore Longitudinal Study and have previously been described in greater detail (8, 10). These healthy, well-educated, community-dwelling men come to the Gerontology Research Center at regular intervals (1-2 yr, depending on age) where they spend 2.5 days. In addition to undergoing complete physical examinations, they are subjects in a variety of physiological, psychological, and clinical studies of aging. Data included in this study were obtained over the period between 1959 and 1975. Subjects with known or suspected thyroid or glucocorticoid dysfunction, as evidenced by history, physical examination, or therapy were excluded from the analyses.

Of this population of subjects, 355 had five or more visits during which both $WB\dot{V}o_2$ and 24-h creatinine excretion had been determined. Methodologies for these measurements were previously described in detail (8, 12).

Nonmuscle $\dot{V}o_2$ ($NM\dot{V}o_2$) and muscle $\dot{V}o_2$ ($M\dot{V}o_2$) in the basal state were calculated from paired values of $WB\dot{V}o_2$ and 24-h creatinine excretions as previously described (12). Each subject's data are represented by a slope obtained by the least-squares linear regression model of each variable on age. They were then grouped by age into decade categories spanning 10 yr, e.g., 15–24, 25–34, 35–44 yr, and so on. In the illustrations each age-decade group is shown by a line centered on the mean value for the variable identified on the ordinate and the group's mean age on the abcissa. The slope of each line represents the mean rate of change of that variable with age; its length along the abcissa represents the group's mean time over which measurements were obtained. Age-related differences in slopes were tested for statistical significance using one-way analyses of variance (BMDP1V).

All calculations were performed by Fortran IV software on a Raytheon digital computer. Data were copied onto digital magnetic tape from IBM tabulation cards that were coded and verified from the laboratory work sheets.

RESULTS AND DISCUSSION

Figure 1 illustrates the mean values and rates of change with age for both the $WB\dot{V}o_2$ (upper panel) and the calculated $NM\dot{V}o_2$ (lower panel) for each of the age-decade groups. There were no statistically significant differences with age in the negative slopes for the $WB\dot{V}o_2$ among the age-decade groups. Each group's mean value is lower than the one preceding it in age and the direction of change fairly well "predicts" the next group's mean value for $WB\dot{V}o_2$. The average slope for all 355 men was −0.82 ml O_2/min per yr (about 5.7 kcal/day or 2,070 kcal/yr), which, for these subjects, represented a mean of 3.7% decline in $WB\dot{V}o_2$ per decade of life. This rate of change is in good agreement with that of 3.22% per decade reported by Keys et al. (5) for their younger group of men studied between ages 21.9 and 41.3 yr. It is higher, however, than the overall 1.2% per decade those authors reported for their older men studied repeatedly between ages 49.8 and 66.8 yr. This difference will be discussed later.

Mean values for $NM\dot{V}o_2$ (Fig. 1, lower panel) showed no statistically significant differences among age groups. Averaged over the 355 subjects, the mean value of 98.2 ± 0.91 (SD) ml/min is, as expected, in excellent agreement with our earlier cross-sectional mean of 99.1 ml/min (12) and indicates that $WB\dot{V}o_2$, as estimated from 24-h creatinine excretion, accounts for all of the decrease in $WB\dot{V}o_2$ with age.

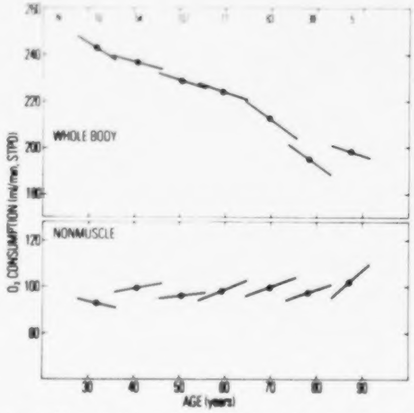


FIG. 1. Mean changes with age in whole-body and nonmuscle O_2 consumption in basal state.

Longitudinal changes in $NM\dot{V}o_2$ are, however, somewhat more perplexing. Except for the youngest group, all other groups exhibit a slight positive slope, suggesting that $NM\dot{V}o_2$ requirements increase with age. However, if this were in fact so, one should expect an overall increase in mean values for this variable; i.e., one group's longitudinal increase should at least tend to "predict" the next group's mean value. This interesting finding led to careful appraisal of factors that are normally not associated with a hypermetabolic state but might account for gradual increases in $\dot{V}o_2$ within the nonmuscle compartment.

The increased prevalence of cardiovascular disorders associated with middle and older ages was of particular interest. Arterial hypertension requires increased work and associated O_2 consumption by the ventricular myocardium. Furthermore, this additional requirement may be disproportionately increased in the presence of concomitant ventricular hypertrophy and the associated decrease in metabolic efficiency of the myocardium (3). Cancer has been associated with a general hypermetabolic state despite decreases in caloric intake (11). Significantly, experimental tranplantation of certain malignant tumors has been associated with definite increases in energy expenditure long before any tumor could be palpated (7).

On this basis a more thorough examination of the subject population for the present study was undertaken. It revealed that of the 355 men included in the analyses, none of which exhibited pathology or conditions generally felt to alter metabolic function, 48 had died since their last measurement. Mean elapsed time since last measurement was 1.9 ± 1.6 (SD) yr. Very few of the 48 decedents had causes of death confirmed by autopsy. It was possible, however, to determine from death certificates, notifications from attending physicians, and review of medical records that 35 of the men's deaths could be attributed primarily to cardiovascular disease and 11 to a variety of malignant processes. Of the remaining two, one died of trauma sustained in a fall and information on the other was unclear. Both were about 80 yr of age and according to our medical records suffered from significant cardiovascular disease. They were, therefore, included in the overall analyses of decedents.

Of the 11 reported to have died of cancer, review of their medical histories revealed 7 who had also been diagnosed as suffering from significant cardiovascular disease such as hypertension, angina, myocardial ischemia (S-T segment depression on ECG at rest), etc. Likewise, 4 of the men whose cause of death was attributed to cardiovascular problems had at one time or another been treated for some malignant disorder (3 skin, 1 multiple myeloma). Thus, 46 of the 48 decedents could each be classified into four groups according to primary and contributing causes of death: cancer alone, cardiovascular alone, cardiovascular with cancer, and cancer with cardiovascular. Mean rates of change with age in $NM\dot{V}o_2$ were compared for these four groups and are shown in Fig. 2. The higher mean values exhibited by those who died as a result of malignant processes with concurrent cardiovascular disorders (CA + CV in Fig. 2), when compared with those for the other groups, suggest that cancer and cardiovascular disease contribute independently to the hypermetabolic state. The small numbers in three of the four groups precluded statistically significant differences with regard to cause of death. (Mean differences between the highest and the lowest slopes approached $P = 0.05$, $t = 2.11$, df = 9.) Therefore, the decedents were grouped by age without regard to cause of death and compared with those presently alive. This is shown in Figs. 3 and 4.

The most striking differences, by age, between the deceased and those alive is the consistent positive slope of $NM\dot{V}o_2$ for those deceased (Fig. 3, lower panel). Examination of the individual points from which slopes were calculated did not suggest a last visit artifact, but a generally consistent and gradual increase in $NM\dot{V}o_2$ throughout the 10-yr period of observation. Except for the two youngest decedents who showed a mean negative slope, there were no statistically significant differences with age among the 48 subjects who had died from all causes. Likewise, no differences with age were found among the slopes of those alive. Averaging the slopes of $NM\dot{V}o_2$ for the 48 dead and the remaining 307 alive men yielded means ± SE of 1.88 ± 0.46 and 0.28 ± 0.23, respectively, for the dead and alive. The difference between the two groups is statistically significant ($t = 2.96$, $P < 0.01$).

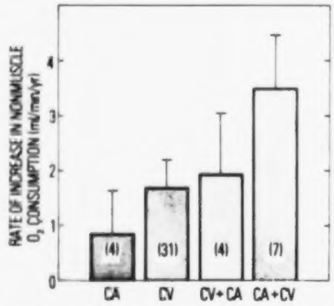


FIG. 2. Means + SE of slopes of nonmuscle O_2 consumption for deceased men grouped by primary cause of death. CA, cancer; CV, cardiovascular. N's for each group are given in parantheses.

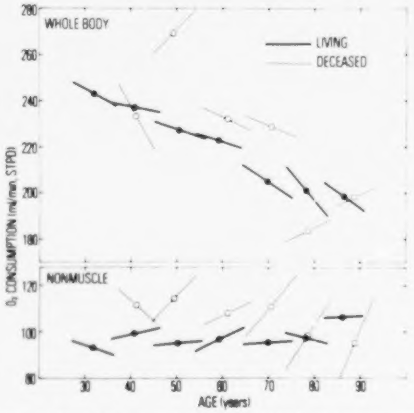


FIG. 3. Mean changes with age in whole-body and nonmuscle consumption for those who died and those still alive (see text).

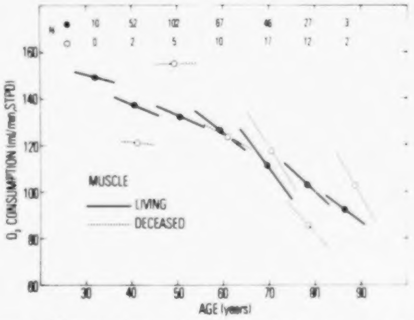


FIG. 4. Mean changes with age in muscle O_2 consumption for those living compared with those who died.

In our earlier cross-sectional study of aging (12) we showed that loss of muscle mass was responsible for all of the decline in $WB\dot{V}o_2$ with age. This conclusion must now be qualified for those nearing terminal age. These men were similar to the survivors in that they lost skeletal muscle at comparable rates (Fig. 4), but differed from the survivors in that they increased their nonmuscle O_2 requirements. Their decline in MVo_2 was offset by gradual increases in NMVo_2 such that WBVo_2 rates of decline were lower, and in our 14 oldest men who were about to die, even reversed (Fig. 3).

Longitudinal surveys of older men, which include only two determinations of WBVo_2 over many years, without the benefit of intermediate measurements in the analyses, might suggest a stabilization of the basal metabolic requirement with age, when in fact there is a reversal in the earlier declining trend. This is evident in the longitudinal data of Keys et al. (5). In their group of 115 men studied five times over a 17-yr period, WBVo_2 decreased from means of 222.8 at age 49.8 to a nadir of 217.7 ml/min (third measurement) at age 54.8 yr. Thereafter, it increased to 220.6 and 218.2 ml/min at ages 60.8 and 66.8 yr, respectively. Thus their conclusion, based on first and last points only, that these older men were "remarkably stable," might, considering the present findings, be interpreted as ominous rather than advantageous.

The present data show that for nearly all of the decade preceding death NMVo_2 requirements gradually increased. We have, therefore, no reliable estimate as to the time of onset of this change in metabolic requirement. It is an unfortunate but reasonable expectation that, as the Baltimore Longitudinal Study continues, more of its participants will die. They will, however, have been observed over longer periods of time so that correlative studies involving increased NMVo_2 and other pertinent physiological and clinical data will be possible.

The autnors are indebted to all participants of the Baltimore Longitudinal Study who have, over the years, given of their time and patience to the study of gerontology. The efforts of those involved in the data collection and recording are also appreciated.

Received 25 January 1978; accepted in final form 23 May 1978.

REFERENCES

1. BEHNKE, A. R. Relationship between basal metabolism, lean body weight and surface area. *Federation Proc.* 12: 13-14, 1953.
2. DUBOIS, D., AND E. F. DUBOIS. A formula to estimate the approximate surface area if height and weight are known. *Arch. Internal Med.* 17: 863-871, 1916.
3. GUNNING, J. R., G. V. COOPER, C. E. HARRISON, AND H. N. COLEMAN. Myocardial oxygen consumption in experimental hypertrophy and congestive heart failure due to pressure overload. *Am. J. Cardiol.* 32: 427-436, 1973.
4. KINNEY, J. M., J. LISTER, AND F. D. MOORE. Relationship of energy expenditure to total exchangeable potassium. In: Body Composition II, edited by J. Brozek. *Ann. N. Y. Acad. Sci.* 110: 711-722, 1963.
5. KEYS, A., H. L. TAYLOR, AND F. GRANDE. Basal metabolism and age of adult man. *Metabolism* 22: 579-587, 1973.
6. MILLER, A. T., JR., AND C. S. BLYTH. Lean body mass as a metabolic reference standard. *J. Appl. Physiol.* 5: 311-316, 1953.
7. PRATT, A. W., AND F. K. PUTNEY. Observations of the energy metabolism of rats receiving Walker tumor 256 transplants. *J. Natl. Cancer Inst.* 20: 173-187, 1958.
8. ROWE, J. W., R. ANDRES, J. D. TOBIN, A. H. NORRIS, AND N. W. SHOCK. The effect of age on creatinine clearance in men: a cross-sectional and longitudinal study. *J. Gerontol.* 31: 155-163, 1976.
9. RUBNER, M. *Die Gesetze des Energieverbrauchs bei der Ernahrung.* Leipzig & Wien: Deutiscke, 1902.
10. STONE, J. L., AND A. H. NORRIS. Activities and attitudes of participants in the Baltimore Longitudinal Study. *J. Gerontol.* 21: 575-580, 1966.
11. THEOLOGIDES, A. Weight loss in cancer patients. *Cancer* 27: 205-208, 1977.
12. TZANKOFF, S. P., AND A. H. NORRIS. Effect of muscle mass decrease on age-related BMR changes. *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.* 43: 1001-1006, 1977.

References

1. Abell, L.L.; Levy, B.B.; Brodie, B.B.; and Kendall, F.E. A simplified method for the estimation of total cholesterol in serum and demonstration of its specificity. *J. Biol. Chem.* *195*:347–366, 1952

2. Abrass, I.B.; and Scarpace, P.J. Human lymphocyte beta-adrenergic receptors are unaltered with age. *J. Gerontol.* *36*: 298–301, 1981

3.[*] Adler, W.H.; Jones, K.H.; and Naruchi, H. Aging and immune function. *In*: Thompson, R.A. (Ed.), Recent Advances in Clinical Immunology. New York: Churchill-Livingstone. 1977

4. Adler, W.H.; and Nagel, J.E. Studies of immune function in a human population. *In*: Segre, D. and Smith, L. (Eds.), Immunological Aspects of Aging. New York and Basel: Marcel Dekker. 1981

5. Agnello, T. Aging and the sense of political powerlessness. *Public Opinion Q.* *37*:251–259, 1973

6. Anderson, J.T.; and Keys, A. Cholesterol in serum and lipoprotein functions—its measurement and stability. *Clin. Chem.* *2*:145–159, 1956

7. Andres, R. Aging and diabetes. *In*: Felig, P. and Bondy, P. (Eds.), Symposium on Diabetes Mellitus. *Med. Clin. North Am.* *55*:835–846, 1971

8. Andres, R. Effect of obesity on total mortality. *Int. J. Obes.* *4*: 381–386, 1980a

9. Andres, R. Influence of obesity on longevity in the aged. *In*: Borek, C., Fenoglio, C.M. and King, D.W. (Eds.), Aging, Cancer and Cell Membranes. New York: Thieme-Stratton. 1980b

10. Andres, R. Aging, diabetes, and obesity: standards of normality. *Mt. Sinai J. Med.* *48*:489–495, 1981

11. Andres, R.; Swerdloff, R.; Pozefsky, T.; and Coleman, D. Manual feedback technique for the control of blood glucose concentration. *In*: Skeggs, L.T., Jr. (Ed.), Chemical Pharmacology. New York: Mediad. 1966

12. Arenberg, D. Anticipation interval and age differences in verbal learning. *J. Abnorm. Psychol.* *70*:419–425, 1965

13. Arenberg, D. Age differences in retroaction. *J. Gerontol.* *22*:88–91, 1967a

14. Arenberg, D. Regression analyses of verbal learning on adult age at two anticipation intervals. *J. Gerontol.* *22*:411–414, 1967b

15. Arenberg, D. Equivalence of information in concept identification. *Psychol. Bull.* *74*:355–361, 1970

16. Arenberg, D. A longitudinal study of problem solving in adults. *J. Gerontol.* *29*:650–658, 1974

17. Arenberg, D. Differences and changes with age in the Benton Visual Retention Test. *J. Gerontol.* *33*:534–540, 1978

18. Arenberg, D. Estimates of age changes on the Benton Visual Retention Test. *J. Gerontol.* *37*:87–90, 1982a

19. Arenberg, D. Changes with age in problem solving. Chapt. 13 *in*: Craik, F.I.M. and Trehub, S. (Eds.), Aging and Cognitive Processes. New York: Plenum. 1982b

20. Arenberg, D. Memory and learning do decline late in life. *In*: Birren, J.E., Munnichs, J.M.A., Thomae, H. and Marois, M. (Eds.), Aging: A Challenge to Science and Society. Vol. 3. Behavioural Sciences and Conclusions. London: Oxford University Press. 1983

---

[*] Italicized numbers indicate studies in the BLSA population.

21. Arenberg, D.; and Robertson-Tchabo, E. Learning and aging. Chapt. 18 *in*: Birren, J.E. and Schaie, K.W. (Eds.), Handbook of the Psychology of Aging. New York: Van Nostrand Reinhold. 1977

22. Asmussen, E.; Fruensgaard, K.; and Nørgaard, S. A follow-up longitudinal study of selected physiological functions in former physical education students—after forty years. *J. Am. Geriatr. Soc. 23*:442–450, 1975

23. Åstrand, I.; Åstrand, P.O.; Hallbäck, I.; and Kilbom, Å. Reduction in maximal oxygen uptake with age. *J. Appl. Physiol. 35*:649–654, 1973

24. Atomic Bomb Casualty Commission. *A General Report on the ABCC-JNIH Joint Research Program, 1947–1975.* Washington, DC: National Academy of Sciences. 1978

25. Baker, H.W.G.; Burger, H.G.; DeKretser, D.M.; Hudson, B.; O'Connor, S.; Wang, C.; Mirovics, A.; Court, J.; Dunlop, M.; and Rennie, G.C.S. Changes in the pituitary testicular system with age. *Clin. Endocrinol. (Oxf.) 5*:349–372, 1976

26. Balke, B.; and Ware, R.W. An experimental study of physical fitness in Air Force personnel. *U.S. Armed Forces Med. J. 10*:675–688, 1959

27. Baltes, P.B. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Hum. Dev. 11*:145–171, 1968

28. Bartoshuk, L.M. The psychophysics of taste. *Am. J. Clin. Nutr. 31*:1068–1077, 1978

29. Baum, B.J. Characteristics of participants in the oral physiology component of the Baltimore Longitudinal Study of Aging. *Community Dent. Oral Epidemiol. 9*:128–134, 1981a

30. Baum, B.J. Evaluation of stimulated parotid saliva flow rate in different age groups. *J. Dent. Res. 60*:1292–1296, 1981b

31. Baum, B.J.; and Bodner, L. Aging and oral motor function: evidence for altered performance among older persons. *J. Dent. Res. 62*:2–6, 1983

32. Baum, B.J.; Kousvelari, E.E.; and Oppenheim, F.G. Exocrine protein secretion from human parotid glands during aging: stable release of the acidic proline-rich proteins. *J. Gerontol. 37*:392–395, 1982

33. Bayley, N.; and Oden, M.H. The maintenance of intellectual ability in gifted adults. *J. Gerontol. 10*:91–107, 1955

34. Becklake, M.R. New index of intrapulmonary mixture of inspired air. *Thorax* 7:111–116, 1952

35. Behnke, A.R. Quantitative assessment of body build. *J. Appl. Physiol. 16*:960–968, 1961

36. Behnke, A.R. Anthropometric evaluation of body composition throughout life. *Ann. NY Acad. Sci. 110*:450–464, 1963

37. Bekesy, G. von. A new audiometer. *Acta Otolaryngol. 35*:411–422, 1947

38. Bell, B. Retinal field shrinkage, age, pulmonary function and biochemistry. *Aging & Hum. Dev. 3*:103–109, 1972a

39. Bell, B. Significance of functional age for interdisciplinary and longitudinal research in aging. *Aging & Hum. Dev. 3*:145–147, 1972b

40. Bell, B.; Rose, C.L.; and Damon, A. The Veterans Administration longitudinal study of healthy aging. *Gerontologist 6*:179–184, 1966

41. Bell, B.; Rose, C.L.; and Damon, A. The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Aging & Hum. Dev. 3*:3–17, 1972

42. Beller, G.A.; Smith, T.W.; Abelmann, W.H.; Haber, E.; and Hood, W.B., Jr. Digitalis intoxication: a prospective clinical study with serum level corrections. *N. Engl. J. Med. 284*:989–997, 1971

43. Belsky, J.L.; Tachikawa, K.; and Jablon, S. The health of atomic bomb survivors: A decade of examinations in a fixed population. *Yale J. Biol. Med. 46*:284–296, 1973

44. Benton, A.L. *The Revised Visual Retention Test: Clinical and Experimental Applications.* 3rd Ed. New York: Psychological Corporation. 1963

45. Berglund, G.; Ander, S.; Lindstrom, B.; and Tibblin, G. Personality and reporting of symptoms in normo- and hypertensive 50-year-old males. *J. Psychosom. Res. 19*:139–145, 1975

46. Berman, M.; and Weiss, M.F. *SAAM Manual.* U.S. Public Health Service Publication No. NIH 76–730. Washington, DC: United States Government Printing Office. 1976

47. Bernreuter, R.G. The theory and construction of the personality inventory. *J. Soc. Psychol. 4*:387–405, 1933

48. Birren, J.E. A neural basis of personal adjustment in aging. *In:* Hanson, P.E. (Ed.), Age with a Future. Copenhagen: Munksgaard. 1964

49. Birren, J.E.; Butler, R.N.; Greenhouse, S.W.; Sokoloff, L.; and Yarrow, M.R. *Human Aging.* U.S. Department of Health, Education and Welfare. National Institute of Mental Health. Publication No. 986. Washington, DC: United States Government Printing Office. 1963

50. Birren, J.E.; and Renner, V.J. Research on the psychology of aging: principles and experimentation. Chapt. 1 *in*: Birren J.E. and Schaie, K.W. (Eds.), Handbook of the Psychology of Aging. New York: Van Nostrand Reinhold. 1977

51. Blankenhorn, D.H.; Rouser, G.; and Weimer, T.J. A method for the estimation of blood glycerides employing florisil. *J. Lipid Res. 2*:281–283, 1961

52. Blomqvist, G. Use of exercise testing for diagnostic and functional evaluation of patients with arteriosclerotic heart disease. *Circulation 44*:1120–1136, 1971

53. Borkan, G.A.; and Norris, A.H. Fat redistribution and the changing body dimensions of the adult male. *Hum. Biol. 49*:495–514, 1977

54. Bosma, J.F. Sensorimotor examination of the mouth and pharynx. *Front. Oral Physiol. 2*:78–107, 1976

55. Botwinick, J. Disinclination to venture response versus cautiousness in responding: age differences. *J. Genet. Psychol. 115*:55–62, 1969

56. Botwinick, J.; and Birren, J.E. A follow-up study of card-sorting performance in elderly men. *J. Gerontol. 20*:208–210, 1965

57. Bowler, R.G.: Determination of thiocyanate in blood serum. *Biochem. J. 38*:385–388, 1944

58. Brandfonbrener, M.; Landowne, M.; and Shock, N.W. Changes in cardiac output with age. *Circulation 12*:557–566, 1955

59. Bregman, E.O. *Revisions of the Army Alpha Examination.* New York: Psychological Corporation. 1925; 1947

60. Brodie, B.B.; Axelrod, J.; Soberman, R.; and Levy, B. The estimation of Antipyrine in biological materials. *J. Biol. Chem. 179*:25–30, 1949

61. Brodman, K.; Erdman, A.J.; and Wolff, G.G. *Cornell Medical Index Health Questionnaire.* New York: Cornell University Medical College. 1949

62. Brodman, K.; Erdman, A.J.; and Wolff, G.G. *The Cornell Medical Index – Health Questionnaire Manual.* New York: Cornell University Medical College. 1960

63. Brozek, J.; and Keys, A. Changes of body weight in normal men who stop smoking cigarettes. *Science 125*:1203, 1957

64. Brückner, R. Longitudinal research on the eye. *Geront. Clin.* 9:87–95, 1967

65. Burney, S.W.; and Bonus, L. Cross-sectional assessment of laboratory variables in a healthy male population. I. Decade age group differences. *Aging & Hum. Dev. 3*:83–88, 1972

66. Busse, E.W. Administration of the interdisciplinary research team. *J. Med. Educ. 40*:832–839, 1965

67. Busse, E.W. A physiological, psychological and sociological study of aging. Chapt. 1 *in*: Palmore, E. (Ed.), Normal Aging. Durham, NC: Duke University Press. 1970

68. Busse, E.W.; and Maddox. G.L. *Final report. The Duke Longitudinal Studies. An Integrated Investigation of Aging and the Aged, Ancillary Studies, and Research Support Services, 1955–1980*. Durham, NC: Duke University Medical Center. Center for the Study of Aging and Human Development. 1980

69. Busse, E.W.; and Maddox, G.L. *The Duke Longitudinal Studies on Aging and the Aged*. New York: Springer. 1983 (In Press)

70. Busse, E.W.; and Obrist, W.D. Significance of focal electroencephalographic changes in the elderly. *In*: Palmore, E. (Ed.), Normal Aging. Durham, NC: Duke University Press. 1970

71. Butler, R.N. Aspects of survival and adaptation in human aging. *Am. J. Psychiatry 123*:1233–1243, 1967

72. Cameron, J.R.; and Sorenson, J.A. Measurement of bone mineral *in vivo*: An improved method. *Science 142*:230–232, 1963

73. Cameron, J.R.; and Sorenson, J.A. Precision and accuracy of bone mineral determination by direct photon absorptiometry. *Invest. Radiol. 3*:141–150, 1968

74. Capriotti, R.; Garwood, M.; and Engel, B.T. Skin potential level: Age and recording site interactions. *J. Gerontol. 36*:40–43, 1981

75. Carpenter, T.M. *Tables, Factors and Formulas for Computing Respiratory Exchange and Biological Transformations of Energy*. 3rd Ed. Washington, DC: Carnegie Institution of Washington. No. 303B. 1939

76. Cattell, R.B. *Personality and Mood by Questionnaire*. San Francisco: Jossey-Bass. 1973

77. Cavan, R.S.; Burgess, E.E.; Havighurst, R.J.; and Goldhamer, H. *Personal Adjustment in Old Age*. Chicago: Science Research Associates. 1949

78. Chabner, B.; and Livingston, D. A simple enzymatic assay for pyridoxal phosphate. *Anal. Biochem. 34*:413–423, 1970

79. Chen, P.S.; Toribara, T.Y.; and Warner, H. Microdetermination of phosphorus. *Anal. Chem. 28*:1756–1758, 1956

80. Chrest, F.J.; Nagel, J.E.; Pyle, R.S.; and Adler, W.H. Human B cell function in responder and nonresponder individuals. II. The role of T helper cells in promoting the PMW-induced B cell production of immunoprotein. *Clin. Exp. Immunol. 53*:465–472, 1983

81. Christensen, P.R.; and Guilford, J.P. *Manual for Christensen-Guilford Fluency Tests*. 2nd Ed. Beverly Hills, CA: Sheridan Psychological Services. 1959

82. Christensen, P.R.; Merrifield, P.R.; and Guilford, J.P. *Consequences: Manual for Administration, Scoring, and Interpretation*. Beverly Hills, CA: Sheridan Psychological Services. 1958

83. Christie, M.J.; and Venables, P.H. Characteristics of palmar skin potential and conductance in relaxed human subjects. *Psychophysiology 8*:525–532, 1971

84. Corso, J.F. Evaluation of operating conditions on a Bekesy-type audiometer. *Arch. Otolaryngol. 61*:649–653, 1955

85. Costa, P.T., Jr.; Fleg, J.L.; McCrae, R.R.; and Lakatta, E.G. Neuroticism, coronary artery disease, and chest pain complaints: Cross-sectional and longitudinal studies. *Exp. Aging Res. 8*:37–44, 1982

86. Costa, P.T., Jr.; and McCrae, R.R. Somatic complaints in males as a function of age and neuroticism: A longitudinal analysis. *J. Behav. Med. 3*:245–255, 1980a

87. Costa, P.T., Jr.; and McCrae, R.R. Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *J. Pers. Soc. Psychol. 38*:668–678, 1980b

88. Costa, P.T., Jr.; and McCrae, R.R. Still stable after all these years: Personality as a key to some issues in adulthood and old age. *In*: Baltes, P.B. and Brim, O.G., Jr. (Eds.), Life-Span Development and Behavior. Vol. 3. New York: Academic Press. 1980c

89. Costa, P.T., Jr.; and McCrae, R.R. Functional age: A conceptual and empirical critique. *In*: Haynes, S.G. and Feinleib, M. (Eds.), Second Conference on the Epidemiology of Aging. NIH Publication No. 80-969. Washington, D.C: United States Government Printing Office. 1980d

90. Costa, P.T., Jr.; and McCrae, R.R. An approach to the attribution of aging, period, and cohort effects. *Psychol. Bull. 92*:238–250, 1982

91. Costa, P.T., Jr.; and McCrae, R.R. Concurrent validation after 20 years: the implications of personality stability for its assessment. *In*: Butcher, J.N. and Spielberger, C.D. (Eds.), Advances in Personality Assessment. Vol. 4. Hillsdale, NJ: Lawrence Erlbaum Assoc. (In Press)

92. Costa, P.T., Jr.; and McCrae, R.R. Personality as a lifelong determinant of well-being. *In*: Malatesta, C. and Izard, C. (Eds.), Affective Processes in Adult Development and Aging. Beverly Hills, Ca: Sage Publications. (In Press)

93. Costa, P.T., Jr.; McCrae, R.R.; Andres, R.; and Tobin, J.D. Hypertension, somatic complaints and personality. Chap. 7 *in*: Elias, M.F. and Streeten, D.H.P. (Eds.), Hypertension and Cognitive Processes. Mt. Desert, ME: Beech Hill. 1980a

94. Costa, P.T., Jr.; McCrae, R.R.; and Arenberg, D. Enduring dispositions in adult males. *J. Pers. Soc. Psychol. 38*:793–800, 1980b

95. Costa, P.T., Jr.; McCrae, R.R.; and Arenberg, D. Recent longitudinal research on personality and aging. Chapt. 7 *in*: Schaie, K.W. (Ed.), Longitudinal Studies of Adult Psychological Development. New York: Guilford. 1983

96. Costa, P.T., Jr.; McCrae, R.R.; and Norris, A.H. Personal adjustment to aging: longitudinal prediction from neuroticism and extraversion. *J. Gerontol. 36*:78–85, 1981

97. Costa, P.T., Jr.; and Shock, N.W. New longitudinal data on the question of whether hypertension influences intellectual performance. Chapt. 6 *in*: Elias, M.F. and Streeten, D.H.P. (Eds.), Hypertension and Cognitive Processes. Mt. Desert, ME: Beech Hill. 1980

98. Cowart, B.J.; and Baum, B.J. Suprathreshold taste sensitivity: changes with post-adolescent development. Presented at the Third Annual Meeting of the Association for Chemoreception Sciences, May 1981

99. Crowne, D.; and Marlowe, D. *The Approval Motive*. New York: Wiley. 1964

100. Cummins, H.; and Midlo, C. *Fingerprints, Palms and Soles*. Philadelphia: Blakiston. 1943

101. Damon, A.; Seltzer, C.C.; Stoudt, H.W.; and Bell, B. Age and physique in healthy white veterans at Boston. *J. Gerontol. 27*:202–208, 1972

102. Darling, R.C.; Cournand, A.; and Richards, D.W., Jr. Studies on intrapulmonary mixture of gases. III. An open circuit method for measuring residual air. *J. Clin. Invest. 19*:609–618, 1940

103. Das, D.N.; Fleg, J.L.; and Lakatta, E.G. Effect of age on the components of atrioventricular conduction in normal man. (Abstract). *Am. J. Cardiol. 49*:(Pt. II), 1031, 1982

104. Davies, D.F., and Shock, N.W. Age changes in glomerular filtration rate, effective renal plasma flow, and tubular excretory capacity in adult males. *J. Clin. Invest. 29*:496–507, 1950

105. Dawber, T.R.; Kannel, W.B.; and Lyell, L.P. An approach to longitudinal studies in a community: the Framingham study. *Ann. NY Acad. Sci. 107*: 539–556, 1963

106. Dawber, T.R.; Meadors, G.F.; and Moore, F.E.; Jr. Epidemiological approaches to heart disease: the Framingham study. *Am. J. Public Health 41*:279–286, 1951

107. Dawber, T.R.; and Thomas, H.E., Jr. Clinical evaluation in the Normative Aging Study. *Aging & Hum. Dev.* *3*:63–69, 1972

108. Dearborn, W.F.; Rothney, J.W.M.; and Shuttleworth, F.K. Data on the growth of public school children. *Monogr. Soc. Res. Child Dev.* *3*:(1), 1–136, 1938

109. DeFronzo, R.A.; Tobin, J.D.; and Andres, R. Glucose clamp technique: a method for quantifying insulin secretion and resistance. *Am. J. Physiol.* *237*:E214–E223, 1979

110. Dehn, M.M.; and Bruce, R.A. Longitudinal variations in maximal oxygen intake with age and activity. *J. Appl. Physiol.* *33*:805–807, 1972

111. Dill, D.B.; Robinson, S.; and Ross, J.C. A longitudinal study of 16 champion runners. *J. Sport Med.* *7*:4–27, 1967

112. Dock, D.S.; and Fukushima, K. A longitudinal study of blood pressure in the Japanese, 1958–1972. *J. Chronic Dis.* *31*:669–689, 1978

113. Douglas, K.; and Arenberg, D. Age changes, cohort differences, and cultural change on the Guilford-Zimmerman Temperament Survey. *J. Gerontol.* *33*:737–747, 1978

114. Drews, R.C. The practice of tonography. Chapt. 6 *in*: Becker, B. and Drews, R.C. (Eds.), Current Concepts in Ophthalmology. St. Louis, MO: Mosby. 1967

115. Edelberg, R. Biopotentials from the skin surface: The hydration effect. *Ann. NY Acad. Sci.* *148*:252–262, 1968

116. Edelman, N.H.; Mittman, C.; Norris, A.H.; Cohen, B.H.; and Shock, N.W. The effects of cigarette smoking upon spirometric performance of community dwelling men. *Am. Rev. Respir. Dis.* *94*:421–429, 1966

117. Edelman, N.H.; Mittman, C.; Norris, A.H.; and Shock, N.W. Effects of respiratory pattern on age differences in ventilation uniformity. *J. Appl. Physiol.* *24*:49–53, 1968

118. Edwards, D.A.W.; Hammond, W.H.; Healy, M.J.R.; Tanner, J.M.; and Whitehouse, R.H. Design and accuracy of calipers for measurement of subcutaneous tissue thickness. *Br. J. Nutr.* 9:133–143, 1955

119. Eichorn, D.H.; Clausen, J.A.; Haan, N.; Honzik, M.P.; and Mussen, P.H. (Eds.) *Present and Past in Middle Life.* New York: Academic Press. 1981

120. Elahi, D.; Potter, J.; Bauman, M.; Tobin, J.; and Andres, R. Interplay of weight, smoking, age and mortality. *In*: Proceedings of the IV International Congress on Obesity. Thorofare, NJ: Slack. (In Press)

121. Elahi, V.K.; Elahi, D.; Andres, A.; Tobin, J.D.; Butler, M.G.; and Norris, A.H. A longitudinal study of nutritional intake in men. *J. Gerontol.* *38*:162–180, 1983

122. Engel, B.T.; and Malmstrom, E.J. An analysis of blood pressure trends based on annual observations of the same subjects. *J. Chronic Dis.* *20*:29–43, 1967

123. Ensor, R.E.; Fleg, J.L.; Kim, Y.C.; deLeon, E.F.; and Goldman, S.M. Longitudinal chest x-ray changes in normal men. *J. Gerontol.* *38*:307–314, 1983

124. Epstein, F.H. An epidemiological study in a total community: the Tecumseh project. *Univ. Mich. Med. Bull.* *206*:307–314, 1960

125. Epstein, F.H.; Francis, T., Jr.; Hayner, N.S.; Johnson, B.C.; Kjelsberg, M.O.; Napier, J.A.; Ostrander, L.D., Jr.; Payne, M.W.; and Dodge, H.J. Prevalence of chronic diseases and distribution of selected physiologic variables in a total community, Tecumseh, Michigan. *Am. J. Epidemiol.* *81*:307–322, 1965

126. Eysenck, H.J. *The Structure of Human Personality.* London: Methuen. 1960

127. Finch, S.C.; and Beebe, G.W. Review of thirty years study of Hiroshima and Nagasaki atomic bomb survivors. II. Biological effects. F. Aging. *J. Radiat. Res. (Tokyo)* *16*:(Suppl.), 108–121, 1975

128. Fisher, M.B.; and Birren, J.E. Age and Strength. *J. Appl. Psychol.* *31*:490–497, 1947

129. Fiske, D.W. The limits for the conventional science of personality. *J. Pers.* *42*:1–11, 1974

130. Fleg, J.L.; Das, D.N.; and Lakatta, E.G. Right bundle branch block: Long-term prognosis in apparently healthy men. *J. Am. Coll. Cardiol. 1*:887–892, 1983a

131. Fleg, J.L.; Gerstenblith, G.; Becker, L.C.; Lakatta, E.G.; and Weisfeldt, M.L. Prognostic value of exercise electrocardiography (EE) and thallium scintigraphy (TS) in asymptomatic subjects. Presented at the American Heart Association's 56th Scientific Sessions, Anaheim, CA, November 1983b

132. Fleg, J.L.; and Kennedy, H.L. Cardiac arrhythmias in a healthy elderly population. *Chest 81*:302–307, 1982

133. Fozard. J.L. Predicting age in the adult years from psychological assessments of abilities and personality. *Aging & Hum. Dev. 3*:175–182, 1972

134. Fozard, J.L.; Nuttall, R.L.; and Waugh, N. Age-related differences in mental performance. *Aging & Hum Dev. 3*:19–43, 1972

135. *Framingham Study: An Epidemiological Investigation of Cardiovascular Disease.* Sect. 1 and 2. Bethesda, MD: U.S. Department of Health, Education and Welfare. National Heart Institute. 1968; Sect. 30. U.S. Department of Health, Education and Welfare Publication No. (NIH) 74–599. Washington, DC: United States Government Printing Office. 1974; Sect. 33. U.S. Department of Health, Education and Welfare Publication No. (NIH) 79–1671. Washington, DC: United States Government Printing Office. 1978

136. Friedlaender, J.S.; Costa, P.T., Jr.; Bosse, R.; Ellis, E.; Rhoads, J.G.; and Stoudt, H.W. Longitudinal physique changes among healthy white veterans at Boston. *Hum. Biol. 49*:541–558, 1977

137. Froelicher, V.F. Use of the exercise electrocardiogram to identify latent coronary atherosclerotic heart disease. *In*: Amsterdam, E.A., Wilmore, J.H. and DeMaria, A.N. (Eds.), Exercise in Cardiovascular Health and Disease. New York: Yorke Medical Books. 1977

138. Garn, S.M. Fat patterning and fat intercorrelations in the adult male. *Hum. Biol. 26*:55–69, 1954

139. Garn, S.M. Radiographic analysis of body composition. *In*: Brozek, J. and Henschel, A. (Eds.), Techniques for Measuring Body Composition. Washington, DC: National Academy of Sciences. 1961

140. Garn, S.M. *The Earlier Gain and the Later Loss of Cortical Bone.* Springfield, IL: Thomas. 1970

141. Garvey, A.J. *Analysis of Eight Variables Common to Six Longitudinal Aging Studies.* Boston: Veterans Administration Out-Patient Clinic. 1973

142. Garwood, M.; Engel, B.T.; and Kusterer, J.P. Skin potential level: Age and epidermal hydration effects. *J. Gerontol. 36*:7–13, 1981

143. Garwood, M.K.; Engel, B.T.; and Quilter, R.E. Age differences in the effect of epidermal hydration on electrodermal activity. *Psychophysiology 16*:311–317, 1979

144. Gelfant, S.; and Grove, G.L. Cycling vs noncycling cells as an explanation for the aging process. *In*: Rockstein, M. (Ed.), Theoretical Aspects of Aging. New York: Academic Press. 1974

145. Gerstenblith, G.; Fleg, J.L.; VanTosh, A.; Becker, L.; Kallman, C.; Andres, R.; Weisfeldt, M.; and Lakatta, E.G. Stress testing redefines the prevalence of coronary artery disease in epidemiologic studies. (Abstract). *Circulation 62*:(4, Pt. II), III-308, 1980

146. Gerstenblith, G.; Frederiksen, J.; Yin, F.C.P.; Fortuin, N.J.; Lakatta, E.G.; and Weisfeldt, M.L. Echocardiographic assessment of a normal adult aging population. *Circulation 56*:273–278, 1977

147. Gerstenblith, G.; Lakatta, E.G.; and Weisfeldt, M.L. Age changes in myocardial function and exercise response. *Prog. Cardiovasc. Dis. 19*:1–21, 1976

148. Giambra, L.M. Daydreaming across the life span: late adolescent to senior citizen. *Int. J. Aging Hum. Dev.* 5:115–140, 1974

149. Giambra, L.M. Daydreaming about the past. The time setting of spontaneous thought intrusions. *Gerontologist* 17:35–38, 1977a

150. Giambra, L.M. A factor analytic study of daydreaming, imaginal process, and temperament: a replication on an adult male life-span sample. *J. Gerontol.* 32:675–680, 1977b

151. Giambra, L.M. Adult male daydreaming across the life span: a replication, further analysis, and tentative norms based upon retrospective reports. *Int. J. Aging Hum. Dev.* 8:197–228, 1977–78

152. Giambra, L.M. Sex differences in daydreaming and related mental activity from the late teens to the early nineties. *Int. J. Aging Hum. Dev.* 10:1–34, 1979–80

153. Giambra, L.M.; and Martin, C.E.M. Sexual daydreams and quantitative aspects of sexual activity: some relations for males across adulthood. *Arch. Sex. Behav.* 6:497–505, 1977

154. Gladis, M.; and Braun, H.W. Age differences in transfer and retroaction as a function of intertask response similarity. *J. Exp. Psychol.* 55:25–30, 1958

155. Gleser, G.; and Ihilevich, D. An objective instrument for measuring defense mechanisms. *J. Consult. Clin. Psychol.* 33:51–60, 1969

156. Gordon, T.; Moore, F.E.; Shurtleff, D.; and Dawber, T.R. Some methodologic problems in the long-term study of cardiovascular disease: observations on the Framingham Study. *J. Chronic Dis.* 10:186–206, 1959

157. Granath, A.; Johnsson, B.; and Strandell, T. Circulation in healthy old men studied by right heart catheterization at rest and during exercise in supine and sitting position. *Acta Med. Scand.* 176:425–446, 1964

158. Granick, S.; and Patterson, R.D. (Eds.) *Human Aging II.* U.S. Department of Health, Education and Welfare. Publication No. (HSM) 71-9037. Washington, DC: United States Government Printing Office. 1971

159. Gsell, O.R. Longitudinal gerontological research over 10 years. (Basel studies, 1955–1965). *Geront. Clin.* 9:67–80, 1967

160. Gsell, O.R. Longitudinale Altersforschung (Präklinische Geriatrie). *Hippokrates* 44:284–293, 1973

161. Guilford, J.P.; and Zimmerman, W.S. *The Guilford-Zimmerman Temperament Survey: Manual of Instructions and Interpretations.* Beverly Hills, CA: Sheridan Supply Company. 1949

162. Guilford, J.P.; and Zimmerman, W.S. Fourteen dimensions of temperament. *Psychol. Monogr.* 70:(417), 1–26, 1956

163. Guilford, J.S.; Zimmerman, W.S.; and Guilford, J.P. *The Guilford-Zimmerman Temperament Survey Handbook.* San Diego, CA: EDITS. 1976

164. Hare, R.S. Endogenous creatinine in serum and urine. *Proc. Soc. Exp. Biol. Med.* 74:148–151, 1950

165. Harlan, W.R.; Graybiel, A.; and Osborne, R.K. Determinants of cardiovascular disease in a young population. *Am. J. Cardiol.* 15:1–12, 1965

166. Harman, D. The aging process. *Proc. Natl. Acad. Sci. USA* 78:7124–7128, 1981

167. Harman, S.M. Aging, sexual activity and sex hormones in healthy men. *Generations* 6:10–13, 1981

168. Harman, S.M.; and Danner, R.L. Rapid measurement of an index of testosterone binding to serum binding globulin using ion exchange columns. *J. Clin. Endocrinol. Metab.* 45:953–959, 1977

169. Harman, S.M.; and Tsitouras, P.D. Reproductive hormones in aging men. I. Measurement of sex steroids, basal luteinizing hormone, and Leydig cell response to human chorionic gonadotropin. *J. Clin. Endocrinol. Metab.* 51:35–40, 1980

170. Harman. S.M.; Tsitouras, P.D.; Costa, P.T.; and Blackman, M.R. Reproductive hormones in aging men. II. Basal pituitary gonadotropins and gonadotropin responses to leuteinizing hormone releasing hormone. *J. Clin. Endocrinol. Metab. 54*:537–541, 1982

171. Harman, S.M.; Tsitouras, P.D.; Kowatch, M.A.; and Kowarski, A.A. Advantage of florisil over charcoal separation in a mechanized testosterone radioimmunoassay. *Clin. Chem. 26*:1613–1616, 1980

172. Hart, M.C.; Orzalese, M.M.; and Cook, C.D. Relation between anatomic dead space and body size and lung volume. *J. Appl. Physiol. 18*:519–522, 1963

173. Heaf, P.J.D.; and Prime, F.J. The compliance of the thorax in normal human subjects. *Clin. Sci 15*:319–327, 1955

174. Helderman, J.H.; Vestal, R.E.; Rowe, J.W.; Tobin, J.D.; Andres, R.; and Robertson, G. The response of arginine vasopressin to intravenous ethanol and hypertonic saline in man: the impact of aging. *J. Gerontol. 33*:39–47, 1978

175. Hershcopf, R.J.; Elahi, D.; Andres, R.; Baldwin, H.L.; Raizes, G.S.; Schocken, D.D.; and Tobin, J.D. Longitudinal changes in serum cholesterol in man: an epidemiologic search for an etiology. *J. Chronic Dis. 35*:101–114, 1982

176. Higgins, M.W.; and Keller, J.B. Seven measures of ventilatory lung function. *Am. Rev. Respir. Dis. 108*:258–272, 1973

177. Hirsh, I.J.: Bekesy's audiometer. *J. Acoust. Soc. Am. 34*:1333–1336, 1962

178. Hollenberg, M.D.; and Schneider, E.L. Receptors for insulin and epidermal growth factor-urogastrone in adult human fibroblasts do not change with donor age. *Mech. Ageing Dev. 11*:37–43, 1979

179. Hollingsworth, J.W.; Hashizume, A.; and Jablon, S. Correlations between tests of aging in Hiroshima subjects. An attempt to define "physiologic age." *Yale J. Biol. Med. 38*:11–26, 1965

180. Hügin, F.; Norris, A.H.; and Shock, N.W. Skin reflex and voluntary reaction times in young and old males. *J. Gerontol. 15*:388–391, 1960

181. Hurst, J.W.; Logue, R.B.; Schlant, R.C.; and Wenger, N.K. (Eds.) *The Heart.* New York: McGraw-Hill. 1978

182. Jenss, R. Age variations of systolic blood pressure in United States Army officers. *Am. J. Hyg. 20*:574–603, 1934

183. Jones, H.E. Problems of method in longitudinal research. *Vita Humana 1*:93–99, 1958

184. Jones, H.E.; and Conrad, H.S. The growth and decline of intelligence; a study of a homogeneous group between the ages of ten and sixty. *Genet. Psychol. Monogr. 13*:223–298, 1933

185. Jones, M.C.; Bayley, N.; MacFarlane, J.W.; and Honzik, M.P. *The Course of Human Development.* Waltham, MA: Xerox College. 1971

186. Kannel, W.B. Recent findings of the Framingham Study. *Resident & Staff Physic. 24*:56–61; 64–66; 71, Jan. 1978

187. Kapur, K.K.; Glass, R.L.; Loftus, E.R.; Alman, J.E.; and Feller, R.P. The Veterans Administration longitudinal study of oral health and disease: methodology and preliminary findings. *Aging Hum. Dev. 3*:125–137, 1972

188. Kelso, L.E.A.; and Hellebrandt, F.A. The recording electrodynamic brake bicycle ergometer. *J. Lab. Clin. Med. 19*:1105–1113, 1934

189. Kenny, D.A. *Correlation and Causality.* New York: Wiley. 1979

190. Kessler, G. Automated techniques in lipid chemistry. *Adv. Clin. Chem. 10*:45–64, 1967

191. Keys, A. Overweight, obesity, coronary heart disease and mortality. *Nutr. Rev. 38*:297–307, 1980

192. Keys, A.; Taylor, H.L.; Blackburn, H.; Brozek, J.; Anderson, J.T.; and Simonson, E. Coronary heart disease among Minnesota business and professional men followed fifteen years. *Circulation 28*:381–395, 1963

193. Keys, A.; Taylor, H.L.; Blackburn, H.; Brozek, J.; Anderson, J.T.; and Simonson, E. Mortality and coronary heart disease among men studied for 23 years. *Arch. Intern. Med. 128*:201–214, 1971

194. Keys, A.; Taylor, H.L.; and Grande, F. Basal metabolism and age of adult man. *Metabolism 22*:579–587, 1973

195. Keys, A.; Taylor, H.L.; Simonson, E.; Blackburn, H.; and Anderson, J.T. The C.V.D. research program of the Laboratory of Physiological Hygiene. *Lancet 81*:291–295, 1961

196. Kinsey, A.C.; Pomeroy, W.B.; and Martin, C.E. *Sexual Behavior in the Human Male.* Philadelphia: Saunders. 1948

197. Klimt, C.R.; Prout, T.E.; Bradley, R.F.; Dolger, H.; Fisher, G.; Gastineau, C.F.; Marks, H.; Meinert, C.L.; Schumacher, O.P.; Cooper, G.R.; Mather, A.; Hainline, A.; and Andres, R. Standardization of the oral glucose tolerance test. *Diabetes 18*:299–310, 1969

198. Kory, R.C.; Callahan, R.; Boren, H.G.; and Syner, J.C. The Veterans Administration-Army cooperative study of pulmonary function. I. Clinical spirometry in normal men. *Am. J. Med. 30*:243–258, 1961

199. Lakatta, E.G. Perspectives on the aged myocardium. *Adv. Exp. Med. Biol. 97*:147–169, 1978

200. Lakatta, E.G. Alterations in the cardiovascular system that occur in advanced age. *Fed. Proc. 38*:163–167, 1979

201. Landmann, R.; Bittiger, H.; and Bühler, F.R. High affinity beta-2–adrenergic receptors in mononuclear leukocytes: similar density in young and old normal subjects. *Life Sci. 29*:1761–1771, 1981

202. Leon, G.R.; Gillum, B.; Gillum, R.; and Gouze, M. Personality stability and change over a 30 year period—middle age to old age. *J. Consult. Clin. Psychol. 23*:245–259, 1979

203. Leto, S.; Yiengst, M.J.; and Barrows, C.H., Jr. The effect of age and protein deprivation on the sulfhydryl content of serum albumin. *J. Gerontol. 25*:4–8, 1970

204. Levitt, M.F.; and Gaudino, M. Measurement of body water compartments. *Am. J. Med.* 9:208–215, 1950

205. Lewis, R.C.; Duval, A.M.; and Iliff, A. Standards for the basal metabolism of children from 2 to 15 years of age, inclusive. *J. Pediatr. 23*:1–18, 1943

206. MacIntyre, N.R. 37 years later...1000 aviators. *Nav. Aviat. News* 9–15, June 1978

207. MacIntyre, N.R.; Mitchell, R.E.; Oberman, A.; Harlan, W.R.; Graybiel, A.; and Johnson, E. Longevity in military pilots: 37–year follow-up of the Navy's "1000 Aviators." *Aviat. Space Environ. Med. 49*:1120–1122, 1978

208. MacIntyre, N.R.; Oberman, A.; Harlan, W.; Mitchell, R.E.; and Graybiel, A. *The Thousand Aviator Study: 1969–1971 Follow-up Program with Distributions and Intercorrelations of Selected Variables.* Monograph 24. Pensacola, FL: Naval Aerospace Medical Research Laboratory. 1979

209. Maddi, S.R. *Personality Theories: A Comparative Analysis.* Homewood, IL: Dorsey Press. 1976

210. Maddox, G.L.; and Douglass, E.B. Self-assessment of health; a longitudinal study of elderly subjects. *J. Health Soc. Behav. 14*:87–93, 1973

211. Maddox, G.L.; and Douglass, E.B. Aging and individual differences; a longitudinal analysis of social psychological and physiological indicators. *J. Gerontol. 29*:555–563, 1974

212. Magladery, J.W.; Teasdall, R.D.; and Norris, A.H. Effect of aging on plantar flexor and superficial abdominal reflexes in man—a clinical and electromyographic study. *J. Gerontol. 13*:282–288, 1958

213. Martin, C.E. Marital and sexual factors in relation to age, disease, and longevity. *In*: Wirt, R.D., Winokur, G. and Roff, M. (Eds.), Life History Research in Psychopathology. Vol. 4. Minneapolis: University of Minnesota Press. 1975

214. Martin, C.E. Sexual activity in the aging male. *In*: Money, J. and Musaph, H. (Eds.), Handbook of Sexology. Amsterdam: Elsevier/North Holland. 1977

215. Martin, C.E. Factors affecting sexual functioning in 60–79-year-old married males. *Arch. Sex. Behav. 10*:399–420, 1981

216. Martin, G.M.; Sprague, C.A.; and Epstein, C.J. Replicative life span of cultured human cells: effects of donor age, tissue and genotype. *Lab. Invest. 23*:86–92, 1970

217. Mason, E.P. Some correlates of self-judgments of the aged. *J. Gerontol. 9*:324–337, 1954

218. Mason, K.; and Mason, W. Some methodological issues in cohort analysis of archival data. *Am. Sociol. Rev. 38*:242–258, 1973

219. Master, A.M.; Nuzie, S.; Brown, R.C.; and Parker, R.C. The electrocardiogram and the "two-step" exercise; a test of cardiac function and coronary insufficiency. *Am. J. Med. Sci. 207*:435–449, 1944

220. Master, A.M.; and Oppenheimer, E.T. A simple exercise tolerance test for circulatory efficiency with standard tables for normal individuals. *Am. J. Med. Sci. 177*:223–242, 1929

221. McCrae, R.R. Age differences in the use of coping mechanisms. *J. Geront. 37*:454–460, 1982a

222. McCrae, R.R. Consensual validation of personality traits: Evidence from self-reports and ratings. *J. Pers. Soc. Psychol. 43*:293–303, 1982b

223. McCrae, R.R.; and Costa, P.T., Jr. Openness to experience an ego level in Loevinger's sentence completion test: dispositional contributions to developmental models of personality. *J. Pers. Soc. Psychol. 39*:1179–1190, 1980

224. McCrae, R.R.; and Costa, P.T., Jr. Aging, the life course, and models of personality. Chapt. 36 *in*: Field, T.M., Huston, A., Quay, H.C., Troll, L. and Finley, G.E. (Eds.), Review of Human Development. New York: Wiley. 1982

225. McCrae, R.R.; Costa, P.T., Jr.; and Arenberg, D. Constancy of adult personality structure in males: longitudinal, cross-sectional and times-of-measurement analyses. *J. Gerontol. 35*:877–883, 1980

226. McGandy, R.B.; Barrows, C.H., Jr.; Spanias, A.; Meredith, A.; Stone, J.L.; and Norris, A.H. Nutrient intakes and energy expenditure in men of different ages. *J. Gerontol. 21*:581–587, 1966

227. McGuire, E.A.; Tobin, J.D.; Berman, M.; and Andres, R. Kinetics of native insulin in diabetic, obese, and aged men. *Diabetes 28*:110–120, 1979

228. Medvedev, Z.A. The nucleic acids in development and aging. *In*: Strehler, B.L. (Ed.), Advances in Gerontological Research. Vol. 1. New York: Academic Press. 1964

229. Meredith, H.V. The rhythm of physical growth: a study of 18 anthropometric measurements on Iowa City white males ranging in age between birth and 18 years. *Univ. Iowa Stud. Child Welf. 11*:(3), 1–128, 1935

230. Metropolitan Life Insurance Company. New weight standards for men and women. *Statist. Bull. Metrop. Life Insur. Co. 40*:1–3, Nov.-Dec. 1959

231. Mischel, W. *Personality and Assessment.* New York: Wiley. 1968

232. Mitchell, R.E. The thousand aviators—thirty-three years later. *In*: Rose, C.L. (Ed.), Collaboration Among Longitudinal Studies. Research Report Series, Publication No. 8. Boston: Veterans Administration Outpatient Clinic. 1976

233. Mittman, C.; Edelman, N.H.; Norris, A.H.; and Shock, N.W. The relationship between chest wall and pulmonary compliance and age. *J. Appl. Physiol. 20*:1211–1216, 1965

234. Monnier, M. Changes in pulse wave velocity with age. Longitudinal gerontological research over 10 years. (Basel studies, 1955–1965). *Geront. Clin.* 9:81–86, 1967

235. Montoye, H.J. *Physical Activity and Health: An Epidemiological Study of an Entire Community.* Englewood Cliffs, NJ: Prentice Hall. 1975

236. Montoye, H.J.; Block, W.D.; and Gayle, R. Maximal oxygen uptake and blood lipids. *J. Chronic Dis. 31*:111–118, 1978

237. Montoye, H.J.; Block, W.D.; Keller, J.B.; and Willis, P.W., III. Glucose tolerance and physical fitness. An epidemiologic study in an entire community. *Eur. J. Applied Physiol. 37*:237–242, 1977

238. Montoye, H.J.; Epstein, F.H.; and Kjelsberg, M.O. The measurement of total body fatness: a study in a total community. *Am. J. Clin. Nutr. 16*:417–427, 1965

239. Nagel, J.E.; Chrest, F.J.; and Adler, W.H. Human B cell function in normal individuals of various age. I. *In vitro* enumeration of pokeweed-induced peripheral blood lymphocyte immunoglobulin-synthesizing cells and the comparison of the results with numbers of peripheral B and T cells, mitogen responses and levels of serum immunoglobulins. *Clin. Exp. Immunol. 44*:646–653, 1981a

240. Nagel, J.E.; Chrest, F.J.; and Adler, W.H. Enumeration of T lymphocyte subsets by monoclonal antibodies in young and aged humans. *J. Immunol. 127*:2086–2088, 1981b

241. Nagel, J.E.; Chrest, F.J.; and Adler, W.H. Mitogenic activity of 12–*0*–tetradecanoyl phorbol–13–acetate on peripheral blood lymphocytes from young and aged adults. *Clin. Exp. Immunol. 49*:217–224, 1982a

242. Nagel, J.E.; Chrest, F.J.; Pyle, R.S.; and Adler, W.H. Monoclonal antibody analysis of T-lymphocyte subsets in young and aged adults. *Immunol. Commun. 12*:223–237, 1983

243. Nagel, J.E.; Collins, G.D.; and Adler, W.H. Spontaneous or natural killer cytotoxicity of K562 erythroleukemic cells in normal patients. *Cancer Res. 41*:2284–2288, 1981c

244. Nagel, J.E.; Pyle, R.S.; Chrest, F.J.; and Adler, W.H. Oxidative metabolism and bactericidal capacity of polymorphonuclear leukocytes from normal young and aged adults. *J. Gerontol. 37*:529–534, 1982b

245. Napier, J.A. Field methods and response rates in the Tecumseh community health study. *Am. J. Public Health 52*:208–216, 1962

246. Napier, J.A.; Johnson, B.C.; and Epstein, F.H. The Tecumseh, Michigan community health study. *In*: Kessler, I.J. and Levin, M.L. (Eds.), The Community as an Epidemiologic Laboratory. Baltimore, MD: Johns Hopkins Press. 1970

247. National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose tolerance. *Diabetes 28*:1039–1057, 1979

248. Nesselroade, S.R.; and Baltes, P.B. Adolescent personality development and historical change, 1970–1972. *Monogr. Soc. Res. Child Dev. 39*:(154), 1–80, 1974

249. Neugarten, B.L. *Personality in Middle and Late Life.* New York: Atherton Press. 1964

250. Neugarten, B.L. Personality and aging. Chapt. 26 *in*: Birren, J.E. and Schaie, K.W. (Eds.), Handbook of the Psychology of Aging. New York: Van Nostrand Reinhold. 1977

251. Norris, A.H.; Lundy, T.; and Shock, N.W. Trends in selected indices of body composition in men between the ages of 30 and 80 years. *Ann. NY Acad. Sci. 110*:623–639, 1963

252. Norris, A.H.; Mittman, C.; and Shock, N.W. Changes in ventilation with age. *In*: Cander, L. and Moyer, J.H. (Eds.), Aging of the Lung; Perspectives. New York: Grune and Stratton. 1964

253. Norris, A.H.; and Shock, N.W. Age changes in ventilatory and metabolic responses to submaximal exercise. *In*: Fourth Congress of the International Association of Gerontology, Vol. II. Fidenza, Italy: Tito Mattioli. 1957

254. Norris, A.H.; and Shock, N.W. The six S's. *Newsletter (Gerontol. Soc.)* 7:7, March 1960

255. Norris, A.H.; Shock, N.W.; and Falzone, J.A., Jr. Relation of lung volumes and maximal breathing capacity to age and socio-economic status. *In*: Blumenthal, H.T. (Ed.), Medical and Clinical Aspects of Aging. New York: Columbia University Press. 1962

256. Norris, A.H.; Shock, N.W.; Landowne, M.; and Falzone, J.A., Jr. Pulmonary function studies: Age differences in lung volumes and bellows function. *J. Gerontol.* *11*:379–387, 1956

257. Norris, A.H.; Shock, N.W.; and Wagman, I.H. Age changes in the maximum conduction velocity of motor fibers of human ulnar nerves. *J. Appl. Physiol.* 5:589–593, 1953

258. Nuttall, R.L. The strategy of functional age research. *Aging & Hum. Dev.* *3*:149–152, 1972

259. Oberman, A.; Lane, N.E.; Harlan, W.R.; Graybiel, A.; and Mitchell, R.E. Trends in systolic blood pressure in the thousand aviator cohort over a twenty-four-year period. *Circulation* *36*:812–822, 1967

260. Oberman, A.; Lane, N.E.; Mitchell, R.E.; and Graybiel, A. *The Thousand Aviator Study. Distributions and Inter-correlations of Selected Variables.* Monograph 12. Pensacola, FL: United States Naval School of Aviation Medicine. 1965a

261. Oberman, A.; Mitchell, R.E.; and Graybiel, A. *The Thousand Aviator Study. Methodology.* Monograph 11. Pensacola, FL: United States Naval School of Aviation Medicine. 1965b

262. Oden, M.H. The fulfillment of promise: 40-year follow-up of the Terman gifted group. *Genet. Psychol. Monogr.* *77*:3–93, 1968

263. Owens, W.A. Age and mental abilities; a longitudinal study. *Genet. Psychol. Monogr.* *48*:3–54, 1953

264. Owens, W.A. Age and mental abilities; a second adult follow-up. *J. Educ. Psychol.* *57*:311–325, 1966

265. Palmore, E. (Ed.) *Normal Aging.* Durham, NC: Duke University Press. 1970.

266. Palmore, E. (Ed.) *Normal Aging II.* Durham, NC: Duke University Press. 1974a

267. Palmore, E. Appendix, A. Design of the adaptation study. *In*: Palmore, E. (Ed.), Normal Aging II. Durham, NC: Duke University Press. 1974b

268. Palmore, E. *Social Patterns in Normal Aging: Findings from the Duke Longitudinal Study.* Durham, NC: Duke University Press. 1981

269. Payne, J.P.; Foster, D.V.; Hill, D.W.; and Wood, D.G.L. Observations on interpretation of blood alcohol levels derived from analysis of urine. *Br. Med. J.* *3*:819–823, 1967

270. Peters, J.P.; and Van Slyke, D.D. *Quantitative Clinical Chemistry. Vol I. Methods.* Baltimore, MD: Williams and Wilkins. 1932

271. Pirke, K.M.; and Doerr, P. Age related changes in free plasma testosterone, dihydrotestosterone, and oestradiol. *Acta Endocrinol. (Copenh.)* *80*:171–178, 1975

272. Plato, C.C. Dermatoglyphics and aging. *J. Gerontol.* *33*:31–38, 1978

273. Plato, C.C.; and Norris, A.H. Osteoarthritis of the hand: age-specific joint-digit prevalence rates. *Am. J. Epidemiol.* *109*:169–180, 1979a

274. Plato, C.C.; and Norris, A.H. Osteoarthritis of the hand: longitudinal studies. *Am. J. Epidemiol.* *110*:740–746, 1979b

275. Plato, C.C.; and Norris, A.H. Bone measurements of the second metacarpal and grip strength. *Hum. Biol.* *52*:131–149, 1980

276. Plato, C.C.; and Purifoy, F.E. Age, sex and bilateral variability in cortical bone loss and measurements of the second metacarpal. *Growth* *46*:100–112, 1982

277. Plato, C.C.; Wood, J.L.; and Norris, A.H. Bilateral asymmetry in bone measurements of the hand and lateral hand dominance. *Am. J. Phys. Anthropol.* *52*:27–31, 1980

278. Potter, J.; Elahi, D.; Tobin, J.D.; and Andres, R. Effect of aging on the cardiothoracic ratio of men *J. Am. Geriatr. Soc.* *30*:404–409, 1982

279. Pozefsky, T.; Colker, J.L.; Langs, H.M.; and Andres, R. The cortisone-glucose tolerance test: The influence of age on performance. *Ann. Intern. Med.* *63*:988–1000, 1965

280. Quilter, R.E.; Giambra, L.M.; and Benson, P.E. Longitudinal age changes in vigilance over an eighteen year interval. *J. Gerontol.* *38*:51–54, 1983

281. Ramfjord, S.P. The periodontal disease index. *J. Periodontol.* *38*:602–610, 1967

282. Ramm, D.; and Gianturco, D.T. Appendix B. Data processing in longitudinal studies. *In*: Palmore, E. (Ed.), Normal Aging II. Durham, NC: Duke University Press. 1974

283. Riley, M.W.; Johnson, M.; and Foner, A. *Aging and Society. Volume 3. A Sociology of Age Stratification*. New York: Russell Sage Foundation. 1972

284. Roach, M.K.; and Creaven, P.J. A micromethod for the determination of acetaldehyde and ethanol in blood. *Clin. Chim. Acta* *21*:275–278, 1968

285. Robertson, G.L.; Mahr, E.A.; Athar, S.; and Sinha, T. Development and clinical application of a new method for the radioimmunoassay of arginine vasopressin in human plasma. *J. Clin. Invest.* *52*:2340–2352, 1973

286. Robertson-Tchabo, E.; and Arenberg, D. Age differences in cognition in healthy educated men: a factor analysis of experimental measures. *Exp. Aging Res.* *2*:75–89, 1976

287. Robertson-Tchabo, E.A.; Arenberg, D.; and Costa, P.T., Jr. Temperamental predictors of longitudinal change in performance on the Benton Revised Visual Retention Test among seventy year old men: an exploratory study. *In*: Hoffmeister, F. and Muller, C. (Eds.), Brain Function in Old Age. Evaluation of Changes and Disorders. Bayer-Symp. VII. New York: Springer-Verlag. 1979

288. Robinson, S.; Dill, D.B.; Tzankoff, S.P.; Wagner, J.A.; and Robinson, R.D. Longitudinal studies of aging in 37 men. *J. Appl. Physiol.* *38*:263–267, 1975

289. Rockstein, M. (Ed.) *Theoretical Aspects of Aging*. New York: Academic Press. 1974

290. Rodeheffer, R.; Gerstenblith, G.; Fleg, J.L.; Lakatta, E.G.; Clulow, J.; Kallman, C.H.; Weisfeldt, M.L.; and Becker, L.C. The impact of age on gated blood pool scans (GBPS) measurements of LV volumes during exercise. (Abstract). *Circulation* *64*:(4), Pt. II, IV–243, 1981

291. Roe, A.; and Siegelman, M. A parent-child relations questionnaire. *Child Dev.* *34*:355–369, 1963

292. Rose, C.L. Representativeness of volunteer subjects in a longitudinal aging study. *Hum. Dev.* *8*:152–156, 1965

293. Rose, C.L. (Ed.) *Collaboration Among Longitudinal Aging Studies. Research Report Series, Publication No. 8.* Boston: Veterans Administration Outpatient Clinic. 1976

294. Rose, C.L. and Bell, B. Selection of geographically stable subjects in longitudinal studies of aging. *J. Am. Geriatr. Soc.* *13*:143–151, 1965

295. Rose, C.S.; Gyorgy, P.; Butler, M.; Andres, R.; Norris, A.H.; Shock, N.W.; Tobin, J.; Brin, M.; and Spiegel, H. Age differences in vitamin B_6 status of 617 men. *Am. J. Clin. Nutr.* *29*:847–853, 1976

296. Rose, G.A.; and Blackburn, H. *Cardiovascular Survey Methods. World Health Organization Monograph. Series No. 56.* Geneva: World Health Organization. 1968

297. Rowe, J.W.; Andres, R.; Tobin, J.D.; Norris, A.H.; and Shock, N.W. Age-adjusted standards for creatinine clearance. *Ann. Intern. Med.* *84*:567–569, 1976a

298. Rowe, J.W.; Andres, R.; Tobin, J.D.; Norris, A.H.; and Shock, N.W. The effect of age on creatinine clearance in men: a cross-sectional and longitudinal study. *J. Gerontol.* *31*:155–163, 1976b

299. Rowe, J.W.; Shock, N.W.; and DeFronzo, R.A. The influence of age on the renal response to water deprivation in man. *Nephron* *17*:270–278, 1976c

300. Rubens, R.; Dhont, M.; and Vermeulen, A. Further studies on Leydig cell function in old age. *J. Clin. Endocrinol. Metab.* *39*:40–45, 1974

301. Rudinger, G.; and Schmitz-Scherzer, R. Sample and methods. *In*: Thomae, H. (Ed.), Patterns of Aging. Findings from the Bonn Longitudinal Study of Aging. Volume 3. Contributions to Human Development. Basel: Karger. 1976

302. Sainsbury, P. Neuroticism and hypertension in an outpatient population. *J. Psychosom. Res.* *12*:261–273, 1960

303. Saxena, Q.B.; Mezey, E.; and Adler, W.H. Regulation of natural killer activity *in vivo*. II. The effect of alcohol consumption on human peripheral blood natural killer activity. *Int. J. Cancer* *26*:413–417, 1980

304. Scammon, R.E. The first seriatum study of human growth. *Am. J. Phys. Anthropol.* *10*:329–336, 1927

305. Schaie, K.W. A general model for the study of developmental problems. *Psychol. Bull.* *64*:92–107, 1965

306. Schaie, K.W. Quasi-experimental research designs in the psychology of aging. Chapt. 2 *in*: Birren, J.E. and Schaie, K.W. (Eds.), Handbook of the Psychology of Aging. New York: Van Nostrand Reinhold. 1977

307. Schaie, K.W.; and Labouvie-Vief, G. Generational versus ontogenetic components of change in adult cognitive functioning; a fourteen-year cross-sequential study. *Dev. Psychol.* *10*:305–320, 1974

308. Schaie, K.W.; and Schaie, J.P. Clinical assessment and aging. Chapt. 29 *in*: Birren, J.E. and Schaie, K.W. (Eds.), Handbook of the Psychology of Aging. New York: Van Nostrand Reinhold. 1977

309. Schlesselman, J.J. Planning a longitudinal study. I. Sample size determination. *J. Chronic Dis.* *26*:553–560, 1973a

310. Schlesselman, J.J. Planning a longitudinal study. II. Frequency of measurement and study duration. *J. Chronic Dis.* *26*:561–570, 1973b

311. Schneider, E.L. Aging and cultured human skin fibroblasts. *J. Invest. Dermatol.* *73*:15–18, 1979

312. Schneider, E.L.; and Mitsui, Y. The relationship between *in vitro* cellular aging and *in vivo* human age. *Proc. Natl. Acad. Sci. USA* *73*:3584–3588, 1976

313. Schocken, D.D.; and Roth, G.S. Reduced β-adrenergic receptor concentrations in ageing man. *Nature* *267*:856–858, 1977

314. Schulsinger, F.; Knop, J.; and Mednick, S.A. (Eds.) *Longitudinal Research. Methods and Uses in Behavioral Science.* Hingham, MA: Kluwer Nijhoff. 1981

315. Shaw, D.J.; Rothbaum, D.A.; Angell, C.S.; and Shock, N.W. The effects of age and blood pressure upon the systolic time intervals in males aged 20–89 years. *J. Gerontol.* *28*:133–139, 1973

316. Shock, N.W. Standard values for basal oxygen consumption in adolescents. *Am J. Dis. Child.* *64*:19–32, 1942

317. Shock, N.W. The effect of menarche on basal physiological functions in girls. *Am. J. Physiol.* *139*:288–292, 1943

318. Shock, N.W. Some physiological aspects of adolescence. *Tex. Rep. Biol. Med.* *4*:289–310, 1946

319. Shock, N.W. The United States Public Health Service—Baltimore City Hospitals Research Section on Gerontology. *J. Gerontol.* *2*:169–170, 1947

320. Shock, N.W. Current trends in biological research on aging. *Jpn. J. Geriatr.* *5*:21–27, 1968

321. Shock, N.W. Energy metabolism, caloric intake and physical activity of the aging. *In:* Carlson, L.A. (Ed.), Nutrition in Old Age. X Symposium Swedish Nutrition Foundation. Uppsala: Almqvist & Wiksell. 1972

322. Shock, N.W. Historical perspectives on aging. *In:* Horvath, S. and Yousef, M. (Eds.), Environmental Physiology: Aging, Heat and Altitude. Amsterdam/New York: Elsevier North Holland. 1980

323. Shock, N.W.; Andres, R.; Norris, A.H.; and Tobin, J.D. Patterns of longitudinal changes in renal function. *In:* Orimo, H., Shimada, K., Iriki, M. and Maeda, D. (Eds.), Recent Advances in Gerontologoy. Proc. XI Int'l. Congr. Gerontol. Amsterdam: Excerpta Medica. 1979

324. Shock, N.W.; and Norris, A.H. Neuromuscular coordination as a factor in age changes in muscular exercise. *In*: Brunner, D. and Jokl, E. (Eds.), Medicine and Sport. Vol. 4. Physical Activity and Aging. Basel/New York: S. Karger. 1970

325. Shock, N.W.; Watkin, D.M.; Yiengst, M.J.; Norris, A.H.; Gaffney, G.W.; Gregerman, R.I.; and Falzone, J.A. Age differences in the water content of the body as related to basal oxygen consumption in males. *J. Gerontol. 18*:1–8, 1963

326. Shock, N.W.; and Yiengst, M.J. Age changes in basal respiratory measurements and metabolism in males. *J. Gerontol. 10*:31–40, 1955

327. Shuttleworth, F.K. Sexual maturation and the physical growth of girls age six to nineteen. *Monogr. Soc. Res. Child Dev. 2*:(5), 1–253, 1937

328. Shuttleworth, F.K. The physical and mental growth of girls and boys age six to nineteen in relation to age at maximum growth. *Monogr. Soc. Res. Child Dev. 4*:(3), 1–290, 1939

329. Singer, J.L. *The Inner World of Daydreaming.* New York: Harper & Row. 1975

330. Singer, J.L.; and Antrobus, J.S. A factor-analytic study of daydreaming and conceptually-related cognitive and personality variables. *Percept. Mot. Skills 17*:187–209, 1963

331. Singer, J.L.; and Antrobus, J.S. Daydreaming, imaginal process and personality: a normative study. *In:* Sheehan, P.W. (Ed.), The Function and Nature of Imagery. New York: Academic Press. 1972

332. Siri, W.E. Apparatus for measuring human body volume. *Rev. Sci. Instrum. 27*:729–738, 1956

333. Siri, W.E. Changes in water, fat and density in adults. *In*: Shock, N.W. (Ed.), Aging Around the World. Biological Aspects of Aging. New York: Columbia University Press. 1962

334. Smith, J.R.; Pereira-Smith, O.M.; and Schneider, E.L. Colony size distributions as a measure of *in vivo* and *in vitro* aging. *Proc. Natl. Acad. Sci. USA 75*: 1353–1356, 1978

335. Soberman, R.; Brodie, B.B.; Levy, B.B.; Axelrod, J.; Hollander, V.; and Steele, J.M. The use of antipyrine in the measurement of total body water in man. *J. Biol. Chem. 179*:31–42, 1949

336. Society of Actuaries. *Build and Blood Pressure Study, 1959, Vol. 1.* Chicago: Society of Actuaries. 1960

337. Society of Actuaries and Association of Life Insurance Medical Directors of America. *Build Study, 1979.* Chicago: Society of Actuaries and Association of Life Insurance Medical Directors of America. 1979

338. Stearns, E.L.; MacDonald, J.A.; Kauffman, B.J.; Lucman, T.S.; Winters, J.S.; and Faiman, C. Declining testis function with age: Hormonal and clinical correlates. *Am. J. Med.* 57:761–766, 1974

339. Stieglitz, E.J. The Section on Gerontology in the National Institutes of Health. *Public Health Rep.* 55:2099–2101, 1940

340. Stone, J.L.; and Norris, A.H. Activities and attitudes of participants in the Baltimore Longitudinal Study. *J. Gerontol. 21*:575–580, 1966

341. Suci, G.J.; Davidoff, M.D.; and Surwillo, W.W. Reaction time as a function of stimulus information and age. *J. Exp. Psychol. 60*:242–244, 1960

342. Surwillo, W.W. Frequency of the "alpha" rhythm, reaction time and age. *Nature 191*:823–824, 1961

343. Surwillo, W.W. The relation of simple response time to brain-wave frequency and the effects of age. *Electroenceph. Clin. Neurophysiol. 15*:105–114, 1963a

344. Surwillo, W.W. The relation of response-time variability to age and the influence of brain wave frequency. *Electroenceph. Clin. Neurophysiol. 15*:1029–1032, 1963b

345. Surwillo, W.W. Some observations on the relation of response speed to frequency of photic stimulation under conditions of EEG synchronization. *Electroenceph. Clin. Neurophysiol. 17*:194–198, 1964a

346. Surwillo, W.W. Age and the perception of short intervals of time. *J. Gerontol. 19*:322–324, 1964b

347. Surwillo, W.W. Level of skin potential in healthy males and the influence of age. *J. Gerontol. 20*: 519–521, 1965

348. Surwillo, W.W. On the relation of latency of alpha attenuation to alpha rhythm frequency and the influence of age. *Electroenceph. Clin. Neurophysiol. 20*:129–132, 1966a

349. Surwillo, W.W. The relation of autonomic activity to age differences in vigilance. *J. Gerontol. 21*:257–260, 1966b

350. Surwillo, W.W.; and Quilter, R.E. Vigilance, age, and response-time. *Am. J. Psychol.* 77:614–620, 1964

351. Surwillo, W.W.; and Quilter, R.E. The influence of age on latency time of involuntary (galvanic skin reflex) and voluntary responses. *J. Gerontol. 20*:173–176, 1965a

352. Surwillo, W.W.; and Quilter, R.E. The relation of frequency of spontaneous skin potential responses to vigilance and to age. *Psychophysiol. 1*:272–276, 1965b

353. Swerdloff, R.S.; Pozefsky, T.; Tobin, J.D.; and Andres, R. Influence of age on the intravenous tolbutamide response test. *Diabetes 16*:161–170, 1967.

354. Tanner, J.M. *Growth at Adolescence.* Oxford: Blackwell. 1955

355. Terman, L.M.; and Oden, M.H. *Genetic Studies of Genius. Vol. IV. The Gifted Child Grows Up.* Stanford, CA: Stanford University Press. 1947

356. Terman, L.M.; and Oden, M.H. *Genetic Studies of Genius. Vol. V. The Gifted Group at Mid-Life.* Stanford, CA: Stanford University Press. 1959

357. Thomae, H. (Ed.) *Patterns of Aging. Findings from the Bonn Longitudinal Study of Aging. Volume 3. Contributions to Human Development.* Basel: Karger. 1976

358. Thurstone, L.L.; and Thurstone, T.G. *Examiner Manual for the SRA Primary Mental Abilities Test.* Chicago: Science Research Associates. 1949

359. Tice, R.R.; Schneider, E.L.; Dram, D.; and Thorne, P. Cytokinetic analysis of the impaired proliferative response of peripheral lymphocytes from aged humans to phytohemagglutinin. *J. Exp. Med. 149*:1029–1041, 1979

360. Titmus Optical Co. *Titmus Optical Company Reference Manual. Professional Vision Tester.* Petersburg, VA: 1959

361. Tobin, J.D. Physiological indices of aging. *In*: Danon, D., Shock, N.W. and Marois, M. (Eds.), Aging: A Challenge to Science and Society. Vol. 1. Biology. New York: Oxford University Press. 1981

362. Tripod, J. Organization of longitudinal research on aging. (Basel studies, 1955–1965.) *Geront. Clin. 9*:96–98, 1967

363. Tsitouras, P.D.; Martin, C.E.; and Harman, S.M. Relationship of serum testosterone to sexual activity in healthy elderly men. *J. Gerontol. 37*:288–293, 1982

364. Tuttle, W.W.; and Wendler, A.J. The construction, calibration and use of an alternating current electrodynamic brake bicycle ergometer. *J. Lab. Clin. Med. 30*:173–183, 1945

365. Tzankoff, S.P.; Fleg, J.L.; Norris, A.H.; and Lakatta, E.G. Age-related increase of plasma catecholamines in response to dynamic exercise in healthy adult men. (Abstract). *Physiologist 23*:(4), 50, 1980

366. Tzankoff, S.P.; and Norris, A.H. Effect of muscle mass decrease on age-related BMR changes. *J. Appl. Physiol. 43*:1001–1006, 1977

367. Tzankoff, S.P.; and Norris, A.H. Longitudinal changes in basal metabolism in man. *J. Appl. Physiol. 45*:536–539, 1978

368. Tzankoff, S.P.; and Norris, A.H. Age-related differences in lactate distribution kinetics following maximal exercise. *Eur. J. Applied Pharmacol. 42*:34–40, 1979

369. VanTosh, A.; Lakatta, E.G.; Fleg, J.L.; Weiss, J.; Kallman, C.; Weisfeldt, M.; and Gerstenblith, G. Ventricular dimensional changes during submaximal exercise: effect of aging in normal man. (Abstract). *Circulation 62*:(4, Pt. II), III–129, 1980

370. Vermeulen, A.; Rubens, R.; and Verdonck, L. Testosterone secretion and metabolism in male senescence. *J. Clin. Endocrinol. Metab. 34*:730–735, 1972

371. Verzar, F. Longitudinal gerontological research in Basel over a period of 10 years (1955–1965) *Geront. Clin.* 9:65–66, 1967

372. Vestal, R.E.; McGuire, E.A.; Tobin, J.D.; Andres, R.; Norris, A.H.; and Mezey, E. Aging and ethanol metabolism. *Clin. Pharmacol. Ther. 21*:343–354, 1977

373. Vestal, R.E.; Norris, A.H.; Tobin, J.D.; Cohen, B.H.; Shock, N.W.; and Andres, R. Antipyrine metabolism in man: influence of age, alcohol, caffeine, and smoking. *Clin. Pharmacol. Ther. 18*:425–432, 1975

374. Wagman, I.H.; and Lesse, J. Maximum conduction velocity of motor fibers of ulner nerve in human subjects of various ages and sizes. *J. Neurophysiol. 15*:235–244, 1952

375. Wechsler, D. *WAIS Manual. Wechsler Adult Intelligence Scale.* New York: Psychological Corporation. 1955

376. Weiffenbach, J.M.; Baum, B.J.; and Burghauser, R. Taste thresholds: quality specific variation with human aging. *J. Gerontol. 37*:372–377, 1982

377. Weissler, A.M.; Harris, W.S.; and Schoenfeld, C.D. Bedside techniques for the evaluation of ventricular function in man. *Am. J. Cardiol. 23*:577–583, 1969

378. Welford, A.T. Psychomotor performance. Chapt. 7 *in*: Birren, J.E. (Ed.), Handbook of Aging and the Individual. Chicago: University Chicago Press. 1959

379. Welford, A.T. Motor performance. Chapt. 19 *in*: Birren, J.E. and Schaie, K.W. (Eds.), Handbook of the Psychology of Aging. New York: Van Nostrand Reinhold. 1977

380. Welford, A.T.; Norris, A.H.; and Shock, N.W. Speed and accuracy of movement and their changes with age. *Acta Psychol. (Amst.) 30*:3–15, 1969

381. Wesson, L.G., Jr. Renal hemodynamics in physiological states. *In*: Wesson, L.G., Jr. (Ed.) Physiology of the Human Kidney. New York: Grune and Stratton. 1969

382. Wilkie, F.L.; and Eisdorfer, C. Intelligence and blood pressure in the aged. *Science 172*:959–962, 1971

383. Wilson, A.J.E.; and Webber, I.L. Attrition in a longitudinal study of an aged population. *Exp. Aging Res. 2*:367–387, 1976

384. Wimer, R.E.; and Wigdor, B.T. Age differences in retention of learning. *J. Gerontol. 13*:291–295, 1958

385. Wybenga, D.R.; Ploggi, V.J.; Dirstine, P.H.; and DiGiorgio, J. A direct manual determination of serum total cholesterol with a single stable reagent. *Clin. Chem. 16*:980–984, 1970

386. Yin, F.C.P.; Raizes, G.S.; Guarnieri, T.; Spurgeon, H.A.; Lakatta, E.G.; Fortuin, N.J.; and Weisfeldt, M.L. Age-associated decrease in ventricular response to hemodynamic stress during beta-adrenergic blockade. *Br. Heart J. 40*:1349–1335, 1978

Index of Authors Cited

A

Abell, L.L. 1[*]
Abelmann, W.H. 42
Abrass, I.B. 2
Adler, W.H. *3*[**],*4*,80,*239,240,241,242,243,244*,303
Agnello, T. 5
Alman, J.E. 187
Ander, S. 45
Anderson, J.T. 6,192,193,195
Andres, R. *7,8,9,10,11,93*,109,*120,121,145,174,175*,197,*227,278,279,295,297,298,323,353,372,373*
Angell, C.S. *315*
Antrobus, J.S. 330,331
Arenberg, D. *12,13,14,15,16,17,18,19,20,21,94,95,113,225,286,287*
Asmussen, E. 22
Åstrand, I. 23
Åstrand, P.O. 23
Athar, S. 285
Atomic Bomb Casualty Commission 24
Axelrod, J. 60,335

B

Baker, H.W.G. 25
Baldwin, H.L. *175*
Balke, B. 26
Baltes, P.B. 27,248
Barrows, C.H., Jr. *203,226*
Bartoshuk, L.M. 28
Baum, B.J. *29,30,31,32*,98,*376*
Bauman, M. *120*
Bayley, N. 33,185
Becker, L. *145*
Becker, L.C. *131,290*
Becklake, M.R. 34
Beebe, G.W. 127
Behnke, A.R. 35,36
Bekesy, G. von 37
Bell, B. 38,39,40,41,101,294
Beller, G.A. 42
Belsky, J.L. 43
Benson, P.E. *280*
Benton, A.L. 44
Berglund, G. 45
Berman, M. 46,*227*
Bernreuter, R.G. 47
Birren, J.E. 48,49,50,56,128
Bittiger, H. 201
Blackburn, H. 192,193,195,296
Blackman, M.R. *170*
Blankenhorn, D.H. 51
Block, W.D. 236,237
Blomqvist, G. 52
Bodner, L. *31*
Bonus, L. 65
Boren, H.G. 198
Borkan, G.A. *53*
Bosma, J.F. 54
Bosse, R. 136
Botwinick, J. 55,56
Bowler, R.G. 57
Bradley, R.F. 197
Brandfonbrener, M. 58
Braun, H.W. 154
Bregman, E.O. 59
Brin, M. *295*
Brodie, B.B. 1,60,335
Brodman, K. 61,62
Brown, R.C. 219
Brozek, J. 63,192,193
Bruce, R.A. 110
Brückner, R. 64
Bühler, F.R. 201
Burger, H.G. 25
Burgess, E.E. 77
Burghauser, R. *376*
Burney, S.W. 65
Busse, E.W. 66,67,68,69,70
Butler, M. *295*
Butler, M.G. *121*
Butler, R.N. 49,71

C

Callahan, R. 198
Cameron, J.R. 72,73
Capriotti, R. *74*
Carpenter, T.M. 75
Cattell, R.B. 76
Cavan, R.S. 77
Chabner, B. 78
Chen, P.S. 79
Chrest, F.J. 80,*239,240,241,242,244*
Christensen, P.R. 81,82
Christie, M.J. 83
Clausen, J.A. 119

---

[*] Numbers are identical with those by which publications cited in References are ordered.

[**] Italicized numbers indicate studies in the BLSA population.

Clulow, J. *290*
Cohen, B.H. *116,373*
Coleman, D. *11*
Colker, J.L. *279*
Collins, G.D. *243*
Conrad, H.S. 184
Cook, C.D. 172
Cooper, G.R. 197
Corso, J.F. 84
Costa, P.T., Jr. *85,86*,87,88,89,*90,91,92,93, 94,95,96,97*,136,*170*,223,*224,225,287*
Cournand, A. 102
Court, J. 25
Cowart, B.J. 98
Creaven, P.J. 284
Crowne, D. 99
Cummins, H. 100

D

Damon, A. 40,41,101
Danner, R.L. *168*
Darling, R.C. 102
Das, D.N. *103,130*
Davidoff, M.D. *341*
Davies, D.F. 104
Dawber, T.R. 105,106,107,156
Dearborn, W.F. 108
DeFronzo, R.A. 109,*299*
Dehn, M.M. 110
DeKretser, D.M. 25
deLeon, E.F. *123*
Dhont, M. 300
DiGiorgio, J. 385
Dill, D.B. 111,288
Dirstine, P.H. 385
Dock, D.S. 112
Dodge, H.J. 125
Doerr, P. 271
Dolger, H. 197
Douglas, K. *113*
Douglass, E.B. 210,211
Dram, D. *359*
Drews, R.C. 114
Dunlop, M. 25
Duval, A.M. 205

E

Edelberg, R. 115
Edelman, N.H. *116,117,233*
Edwards, D.A.W. 118
Eichorn, D.H. 119
Eisdorfer, C. 382
Elahi, D. *120,121,175,278*
Elahi, V.K. *121*
Ellis, E. 136
Engel, B.T. *74*,122,*142,143*
Ensor, R.E. *123*
Epstein, C.J. 216
Epstein, F.H. 124,125,238,246
Erdman, A.J. 61,62
Eysenck, H.J. 126

F

Faiman, C. 338
Falzone, J.A. *325*
Falzone, J.A., Jr. *225*,256
Feller, R.P. 187
Finch, S.C. 127
Fisher, G. 197
Fisher, M.B. 128
Fiske, D.W. 129
Fleg, J.L. *85,103,123,130,131,132,145,290, 365,369*
Foner, A. 283
Fortuin, N.J. *146,386*
Foster, D.V. 269
Fozard, J.L. 133,134
Framingham Study: An Epidemiological Investigation of Cardiovascular Disease 135
Francis, T., Jr. 125
Frederiksen, J. *146*
Friedlaender, J.S. 136
Froelicher, V.F. 137
Fruensgaard, K. 22
Fukushima, K. 112

G

Gaffney, G.W. *325*
Garn, S.M. 138,139,140
Garvey, A.J. 141
Garwood, M. *74,142*
Garwood, M.K. *143*
Gastineau, C.F. 197
Gaudino, M. 204
Gayle, R. 236
Gelfant, S. 144
Gerstenblith, G. *131,145,146,147,290,369*
Giambra, L.M. *148,149,150,151,152,153, 280*
Gianturco, D.T. 282
Gillum, B. 202
Gillum, R. 202
Gladis, M. 154
Glass, R.L. 187
Gleser, G. 155
Goldhamer, H. 77
Goldman, S.M. *123*
Gordon, T. 156
Gouze, M. 202
Granath, A. 157
Grande, F. 194
Granick, S. 158

Graybiel, A. 165,207,208,259,260,261
Greenhouse, S.W. 49
Gregerman, R.I. *325*
Grove, G.L. 144
Gsell, O.R. 159,160
Guarnieri, T. *386*
Guilford, J.P. 81,82,161,162,163
Guilford, J.S. 163
Gyorgy, P. *295*

H

Haan, N. 119
Haber, E. 42
Hainline, A. 197
Hallbäck, I. 23
Hammond, W.H. 118
Hare, R.S. 164
Harlan, W. 208
Harlan, W.R. 165,207,259
Harman, D. 166
Harman, S.M. *167,168,169,170,171,363*
Harris, W.S. 377
Hart, M.C. 172
Hashizume, A. 179
Havighurst, R.J. 77
Hayner, N.S. 125
Heaf, P.J.D. 173
Healy, M.J.R. 118
Helderman, J.H. *174*
Hellebrandt, F.A. 188
Hershcopf, R.J. *175*
Higgins, M.W. 176
Hill, D.W. 269
Hirsh, I.J. 177
Hollander, V. 335
Hollenberg, M.D. *178*
Hollingsworth, J.W. 179
Honzik, M.P. 119,185
Hood, W.B., Jr. 42
Hudson, B. 25
Hügin, F. 180
Hurst, J.W. 181

I

Ihilevich, D. 155
Iliff, A. 205

J

Jablon, S. 43,179
Jenss, R. 182
Johnson, B.C. 125,246
Johnson, E. 207
Johnson, M. 283
Johnsson, B. 157
Jones, H.E. 183,184
Jones, K.H. *3*
Jones, M.C. 185

K

Kallman, C. *145,369*
Kallman, C.H. *290*
Kannel, W.B. 105,186
Kapur, K.K. 187
Kauffman, B.J. 338
Keller, J.B. 176,237
Kelso, L.E.A. 188
Kendall, F.E. 1
Kennedy, H.L. *132*
Kenny, D.A. 189
Kessler, G. 190
Keys, A. 6,63,191,192,193,194,195
Kilbom, Å 23
Kim, Y.C. *123*
Kinsey, A.C. 196
Kjelsberg, M.O. 125,238
Klimt, C.R. 197
Knop, J. 314
Kory, R.C. 198
Kousvelari, E.E. *32*
Kowarski, A.A. *171*
Kowatch, M.A. *171*
Kusterer, J.P. *142*

L

Labouvie-Vief, G. 307
Lakatta, E.G. *85,103,130,131,145,146,147,199,200,290,365,369,386*
Landmann, R. 201
Landowne, M. 58,256
Lane, N.E. 259,260
Langs, H.M. *279*
Leon, G.R. 202
Lesse, J. 374
Leto, S. *203*
Levitt, M.F. 204
Levy, B. 60
Levy, B.B. 1,335
Lewis, R.C. 205
Lindstrom, B. 45
Livingston, D. 78
Loftus, E.R. 187
Logue, R.B. 181
Lucman, T.S. 338
Lundy, T. *251*
Lyell, L.P. 105

M

MacDonald, J.A. 338
MacFarlane, J.W. 185
MacIntyre, N.R. 206,207,208
Maddi, S.R. 209

Maddox, G.L. 68,69,210,211
Magladery, J.W. 212
Mahr. E.A. 285
Malmstrom, E.J. 122
Marks, H. 197
Marlowe, D. 99
Martin, C.E. 196,*213,214,215,363*
Martin, C.E.M. *153*
Martin, G.M. 216
Mason, E.P. 217
Mason, K. 218
Mason, W. 218
Master, A.M. 219,220
Mather, A. 197
McCrae, R.R. *85,86*,87,88,89,*90,91,92,93,94,95,96,221,222,223,224,225*
McGandy, R.B. *226*
McGuire, E.A. *227,372*
Meadors, G.F. 106
Mednick, S.A. 314
Medvedev, Z.A. 228
Meinert, C.L. 197
Meredith, A. *226*,
Meredith, H.V. 229
Merrifield, P.R. 82
Metropolitan Life Insurance Company 230
Mezey, E. 303,*372*
Midlo, C. 100
Mirovics, A. 25
Mischel, W. 231
Mitchell, R.E. 207,208,232,259,260,261
Mittman, C. *116,117,233,252*
Mitsui, Y. *312*
Monnier, M. 234
Montoye, H.J. 235,236,237,238
Moore, F.E. 156
Moore, F.E., Jr. 106
Mussen, P.H. 119

N

Nagel, J.E. *4*,80,*239,240,241,242,243,244*
Napier, J.A. 125,245,246
Naruchi, H. *3*
National Diabetes Data Group 247
Nesselroade, S.R. 248
Neugarten, B.L. 249,250
Nørgaard, S. 22
Norris, A.H. *53,96,116,117,121*,180,212,*226,233,251,252*,253,*254,255*,256,257,*273,274,275,277,295,297,298,323,324,325,340,365,366,367,368,372,373,380*
Nuttall, R.L. 134,258
Nuzie, S. 219

O

Oberman, A. 207,208,259,260,261
Obrist, W.D. 70
O'Connor, S. 25
Oden, M.H. 33,262,355,356
Oppenheim, F.G. *32*
Oppenheimer, E.T. 220
Orzalese, M.M. 172
Osborne, R.K. 165
Ostrander, L.D., Jr. 125
Owens, W.A. 263,264

P

Palmore, E. 265,266,267,268
Parker, R.C. 219
Patterson, R.D. 158
Payne, J.P. 269
Payne, M.W. 125
Pereira-Smith, O.M. *334*
Peters, J.P. 270
Pirke, K.M. 271
Plato, C.C. *272,273,274,275,276,277*
Ploggi, V.J. 385
Pomeroy, W.B. 196
Potter, J. *120,278*
Pozefsky, T. *11,279,353*
Prime, F.J. 173
Prout, T.E. 197
Purifoy, F.E. *276*
Pyle, R.S. 80,*242,244*

Q

Quilter, R.E. *143,280,350,351,352*

R

Raizes, G.S. *175,386*
Ramfjord, S.P. 281
Ramm, D. 282
Renner, V.J. 50
Rennie, G.C.S. 25
Rhoads, J.G. 136
Richards, D.W., Jr. 102
Riley, M.W. 283
Roach, M.K. 284
Robertson, G. *174*
Robertson, G.L. 285
Robertson-Tchabo, E. *21,286*
Robertson-Tchabo, E.A. *287*
Robinson, R.D. 288
Robinson, S. 111,288
Rockstein, M. 289
Rodeheffer, R. *290*
Roe, A. 291
Rose, C.L. 40,41,292,293,294
Rose, C.S. *295*
Rose, G.A. 296
Ross, J.C. 111
Roth, G.S. *313*

Rothbaum, D.A. *315*
Rothney, J.W.M. 108
Rouser, G. 51
Rowe, J.W. *174,297,298,299*
Rubens, R. 300,370
Rudinger, G. 301

S

Sainsbury, P. 302
Saxena, Q.B. 303
Scammon, R.E. 304
Scarpace, P.J. 2
Schaie, J.P. 308
Schaie, K.W. 305,306,307,308
Schlant, R.C. 181
Schlesselman, J.J. *309,310*
Schmitz-Scherzer, R. 301
Schneider, E.L. *178,311,312,334,359*
Schocken, D.D. *175,313*
Schoenfeld, C.D. 377
Schulsinger, F. 314
Schumacher, O.P. 197
Seltzer, C.C. 101
Shaw, D.J. *315*
Shock, N.W. 58,*97*,104,*116,117*,180,*233,251,252*,253,*254,255*,256,257,*295,297,298,299,315*,316,317,318,319,*320,321,322,323,324,325*,326,*373,380*
Shurtleff, D. 156
Shuttleworth, F.K. 108,327,328
Siegelman, M. 291
Simonson, E. 192,193,195
Singer, J.L. 329,330,331
Sinha, T. 285
Siri, W.E. 332,333
Smith, J.R. *334*
Smith, T.W. 42
Soberman, R. 60,335
Society of Actuaries 336
Society of Actuaries and Association of Life Insurance Medical Directors of America 337
Sokoloff, L. 49
Sorenson, J.A. 72,73
Spanias, A. *226*
Spiegel, H. *295*
Sprague, C.A. 216
Spurgeon, H.A. *386*
Stearns, E.L. 338
Steele, J.M. 335
Stieglitz, E.J. 339
Stone, J.L. *226,340*
Stoudt, H.W. 101,136
Strandell, T. 157
Suci, G.J. *341*
Surwillo, W.W. *341,342,343,344,345,346,347,348,349,350,351,352*
Swerdloff, R. *11*
Swerdloff, R.S. *353*
Syner, J.C. 198

T

Tachikawa, K. 43
Tanner, J.M. 118,354
Taylor, H.L. 192,193,194,195
Teasdall, R.D. 212
Terman, L.M. 355,356
Thomae, H. 357
Thomas, H.E., Jr. 107
Thorne, P. *359*
Thurstone, L.L. 358
Thurstone, T.G. 358
Tibblin, G. 45
Tice, R.R. *359*
Titmus Optical Co. 360
Tobin, J. *120,295*
Tobin, J.D. *93*,109,*121,174,175,227,278,297,298,323,353,361,372,373*
Toribara, T.Y. 79
Tripod, J. 362
Tsitouras, P.D. *169,170,171,363*
Tuttle, W.W. 364
Tzankoff, S.P. 288,*365,366,367,368*

V

Van Slyke, D.D. 270
VanTosh, A. *145,369*
Venables, P.H. 83
Verdonck, L. 370
Vermeulen, A. 300,370
Verzar, F. 371
Vestal, R.E. *174,372,373*

W

Wagman, I.H. 257,374
Wagner, J.A. 288
Wang, C. 25
Ware, R.W. 26
Warner, H. 79
Watkin, D.M. *325*
Waugh, N. 134
Webber, I.L. 383
Wechsler, D. 375
Weiffenbach, J.M. *376*
Weimer, T.J. 51
Weisfeldt, M. *145,369*
Weisfeldt, M.L. *131,146,147,290,386*
Weiss, J. *369*
Weiss, M.F. 46
Weissler, A.M. 377
Welford, A.T. 378,379,*380*
Wendler, A.J. 364
Wenger, N.K. 181

Wesson, L.G., Jr. 381
Whitehouse, R.H. 118
Wigdor, B.T. 384
Wilkie, F.L. 382
Willis, P.W., III. 237
Wilson, A.J.E. 383
Wimer, R.E. 384
Winters, J.S. 338
Wolff, G.G. 61,62
Wood, D.G.L. 269
Wood, J.L. *277*
Wybenga, D.R. 385

Y

Yarrow, M.R. 49
Yiengst, M.J. *203,325*,326
Yin, F.C.P. *146,386*

Z

Zimmerman, W.S. 161,162,163

Index

A

Acord, Eloise vii
Activities and attitudes 11, 92-93, 166, 197-98
Adaptational outcomes in stress 93
Adjustment to aging 197-98
Administration 47-48
Age-course 208; *App*.292
Aging
  adjustment to 197-98; *App*.215
  "general aging process:" see "'Functional' Age"
  individuality 1, 26, 33, 169-71, 207-208
  in vivo and *in vitro* 96-97
  methods of study 5-18; *App*.92, 354, 362
  patterns of change 208-209
  vs. disease 1-2, 9, 57-60, 114-18, 208-209
  vs. socioeconomic influences 1, 105, 114-15, 135, 156-57, 166, 177-78, 209-10
Aging Cell Repository 88
Alcohol: see "Ethanol"
Alkaloids 58
Alpha-rhythm 153-54
Alveolar collapse 116
Ambulatory ECG 84, 126-27
Amphetamines 57
Analytic design-strategies 10, 15-16, 169-71, 202; *App*.92, 354, 362
Angina pectoris 80-81, 200-202, 204-205
Androgens 135
Angiography 83
Anthropometry 20-21, 24-26, 30-33, 38, 56, 70-72, 97-105
Anti-arrhythmics 57
Antihistamines 58
Antihypertensives 57
Anti-inflammatory agents 57
Antipyrine metabolism 71, 86, 97-99, 112-13, 172
  effect of ethanol, caffeine, smoking 112-13
Aortic valvular stenosis 58
Arginine vasopressin 86-87, 135
Army Alpha test 39-40, 89-90, 185-87, 191
Arteriosclerosis 31
Atomic Bomb Casualty Commission (ABCC) Study 36-39
Attrition 52-55
Audiometry 25, 38, 74, 89
Autonomic activity 124-25, 155
  and cardiovascular function 124-25
  and vigilance 155

B

Balke exercise protocol 77
Ballistocardiogram 20
Baltimore City Hospitals (BCH) vii, 45, 63-80
  Chronic-Disease Hospital vii
  Infirmary Division 45, 115-16, 156-57
  Institutional Review Board 49
Baltimore Longitudinal Study of Aging: see "BLSA"
Basal metabolism (BMR) 11-13, 20-21, 56, 72, 97-99, 110, 174-75; *App*.372, 396
Basel Study 30-32
Behavioral Rigidity test 39
Behnke index 71
Bell, Dr. B. 24
Benton Visual Retention Test 90, 171, 183-85; *App*.14, 21, 335
Berkeley Study 20
Berliner, Dr. Robert vii
Birren, Dr. J.E. 28
Birth-cohort effects 4-5, 8-9, 15-16, 26, 40, 47, 164, 169-71, 187
Blood-chemistry 56
Blood-flow, peripheral 84
Blood-lactate after exercise 40, 83, 99-101
Blood-morphology 38
Blood-pressure 3, 9, 26-27, 31-32, 34, 38-39, 40, 56, 57-59, 72, 77, 83
  and body-weight 27
  and intelligence 185-87; *App*.223
  as predictor of CAD 21
Blood-glucose 25
Blood-type 25, 70
BLSA: conclusions and future directions 207-10
  future directions 209-10
  individuality of aging process 207-208
  patterns of change 208-209
    compensatory change 208-209
    cultural change 208
    declines 207-209
    disease 208-209
    in regulatory mechanisms 209
    no uniform age course 208-209
    precipitous change 208
    stability 208
BLSA: cross-sectional studies 95-167
  behavior 148-67
    activities and attitudes 166
    cognitive functions 158-62
      equivalence of information in concept-identification 161-62

[BLSA: cross-sectional studies (cont'd)]
    learning and memory 158-61
      memory and decision-making 161
      retroaction 161
      verbal learning 158-61
  daydreaming 162-63
    about past events 162
    age differences 162
    relation to temperament 162-63
    sex differences 163
    sexual daydreams 163
  marital and sexual history 164-65
  psychophysiology 148-58
    EEG 151-54
    electrodermal activity 154-56
      autonomic activity and vigilance 155
      epidermal hydration, effect on skin-potential and vigilance 155-56
      latency of involuntary and voluntary responses 154-55
      skin-potential level 155
      skin-potential response and vigilance 154
    reaction-time and vigilance 148-51
    time-perception and psychomotor function 156-58
      institutionalization and age, effect on perception of short time-intervals 156-57
      muscle-strength and coordination 157-58
      speed and accuracy of movement 157
  stress and coping 166-67
interdisciplinary studies 95
physiology 95-148
  body-composition 97-105
    blood-lactate after exercise 99-101
    body-dimensions and fat 101-103
    bone-alterations 103-105
    metabolism and 97-99
  cardiovascular system 117-27
    electrophysiology 126-27
      ambulatory ECG 126-27
      His-bundle ECG, non-invasive assessment 126
    function at rest 118-21, 125-26
      echocardiography, one-dimensional 118-19
      echocardiography, two-dimensional 119
      multiple-gated cardiac blood-pool scan 119-20
      systolic time-intervals 118
      ventricular hypertrophy 120-21
    occult disease 117-18
    response to stress 121-25
      autonomic modulation 124-25
      ejection-fraction 124
      end-diastolic filling volume 123-24
      maximum heart-rate after exercise 122-23
  cell-culture senescence, *in-vitro* life-span 95-97
  drug-metabolism 112-14
    antipyrine 112-13
    ethanol 113-14
  endocrine system 128-37
    arginine vasopressin 135
    diabetes 128-33
    hormone receptors 135-37
    hypothalamic-pituitary-testicular axis 134-35
  immune system 137-47
    antibody-forming ability of peripheral-blood lymphocyte 141-42
    granulocyte function 140-41
    mitogen responses of peripheral blood-cells 146
    natural-killer-cell activity 142-43
    serum-immunoprotein levels 147
    white cells in peripheral blood 137-40
  metabolism and body-composition 97-99
  nutrition 105-12
    caloric intake–diet 105-10
    serum albumin 111-12
    vitamin-B_6 110-11
  oral-dental 148
    gustatory function 148
    motorfunction 148
    salivary-gland function 148
  pulmonary system 114-17
    age differences 114-16
    cigarette-smoking, effects 116-17
  renal system 127-28
    concentrating ability 128
    creatinine clearance 127-28
BLSA: design and operation 45-62
  data-management 60-62
  history 45-46
  interpretation of data 57-60
    aging vs. disease 1-2, 9, 57-60, 114-18, 208-209
    clinical "clean-up" 59-60
  objectives 46
  operation 47-57
    administration 47-48
  participants v-vii, 1, 3, 48-55
    characteristics 51-52
    exclusion from tests 48
    recruitment and attrition 48-55
    sample-size 170
    stability of sample, drop-outs 3, 52-55
    women 1, 56, 103-104, 128, 143, 148, 163, 166-67, 191-92, 207, 209-10
  statistical design 10, 15-16, 169-71, 202; *App.*92, 354, 362

- BLSA: longitudinal studies 169-202; *App.*1-399
  - body-size and composition 171-74; *App.*51, 372
    - fat-redistribution 171-73; *App.*51
    - height, weight 169-70, 173; *App.*372
    - osteoarthritis of hand 173-74; *App.*318
  - cognitive performance 179-87
    - blood-pressure and intelligence 185-87; *App.*223
    - intelligence tests 185-87
    - learning 182-83; *App.*40
    - problem-solving 180-82; *App.*5, 25
    - vigilance 179-80; *App.*331
    - visual memory 183-85; *App.*14, 21, 40
      - and temperament 184-85; *App.*335
  - conceptual approaches 169-71; *App.*92, 354, 362
  - nutrition 171; *App.*245, 372
  - personality 187-202; *App.*171
    - adjustment to aging 197-98; *App.*215
    - historical effects on test-data 192-94
    - hypertension, somatic complaints, and CAD 199-202; *App.*71
      - CAD and personality 200-202; *App.*71
      - hypertension, hypochondriasis, neuroticism 199-200; *App.*147
    - personality and life-course 202; *App.*234, 292
    - response-bias 194-96
    - somatic complaints and neuroticism 198-99; *App.*79
    - stability of personality 187-92; *App.*105, 163, 304
      - definitions, statistical 187-89
      - of factor-structure 192-94
      - of group mean level 189
      - of individual rank-order 189-92
    - stability-model, practical implications 196
  - physiology 174-79
    - basal metabolism 174-75; *App.*372, 396
    - cardiovascular function 177-79
      - CAD and neuroticism *App.*71
      - cardiopulmonary x-ray changes 178; *App.*264
      - cardiothoracic ratio 179; *App.*325
      - right-bundle branch-block 177-78; *App.*272
    - pulmonary function 170; *App.*311
    - renal function 175-76; *App.*345, 384
    - serum cholesterol 176-77; *App.*245, 278
- BLSA: outcome studies 10-13, 95, 202-205; *App.*272, 387
  - coronary events in asymptomatic subjects, prediction 204-205; *App.*272
  - obesity and longevity 202-203
  - physiologic variables 203-204; *App.*387
- BLSA studies: see "BLSA: cross-sectional studies;" "BLSA: longitudinal studies;" "BLSA: outcome studies"
- BLSA: tests administered 55-57; 63-94
  - operation 55-57
    - changes in methods 57
    - criteria of selection 55
    - introduction and elimination 57
    - quality-control 57
    - schedule 55-56
  - tests 1, 63-94
    - body-structure and composition 70-72
      - anthropometry 70
      - basal metabolism 72
      - body-mass 71-72
        - Behnke index 71
        - body-density and fat 71-72
        - body-water compartments, muscle, fat 71
        - muscle-mass 71
      - bone-mineral measurement 70-71
        - bone-density by x-ray 70
        - bone-density by photon scan 70-71
      - subcutaneous fat 71
    - carbohydrate metabolism 84-86
      - cortisone-glucose-tolerance test 85
      - glucose-clamp test 86
      - glucose-tolerance test, intravenous 84-85
      - glucose-tolerance test, oral 85
      - insulin-tolerance test, intravenous 84-85
      - tolbutamide-response test, intravenous 86
    - cardiovascular function 80-84
      - cardiovascular function, effect of age 82-84
        - blood-flow, peripheral 84
        - catecholamine secretion 83
        - digitalis, inotropic effect 84
        - ECG and blood-pressure, ambulatory monitor 84
        - echocardiography 82-83
        - heart-rate response to β-adrenergic stimulation 83
        - His-bundle ECG 83-84
        - multiple-gated cardiac blood-pool (MUGA) scan 83
        - systolic time-intervals 82
        - ventricular response to afterload stress 83
      - CVD screening 80-82
        - CVD history and physical examination 80-81
        - exercise-stress test (double Master) 81-82
        - stress-thallium scans 82
        - treadmill-stress ECG 82

[BLSA: tests administered (cont'd)]
- cell replication 88
- clinical evaluation 63-70
  - Cornell Medical Index 69-70
  - history 63
  - laboratory tests 69
    - cholesterol 69
    - hemogram 69
    - oral-dental 69
    - protein, albumin, and globulin in serum 69
    - syphilis 69
    - triglycerides 69
    - urinalysis 69
  - physical examination 63
- cognitive performance 89-91
  - intelligence tests 89-90
    - Army Alpha test 89-90, 185-87, 191
    - Southern-California tests of mental ability 90
    - vocabulary (WAIS) 90
  - learning, memory, and decision tasks 90-91
    - Benton Visual Retention Test 90
    - memory and decision 90-91
    - verbal learning 90
  - problem-solving 91
    - concept-identification 91
    - logical 91
- drug metabolism 86-87
  - antipyrine 86
  - ethanol 86-87
- genetic characteristics 70
  - family history 70
  - genetic markers 70
    - blood-type 70
    - dermatoglyphics of hand 70
    - hand, foot, eye preference 70
    - taste-test 70
- hypothalamic-pituitary function 87
  - hypothalamic-posterior-pituitary function 87
    - ethanol-infusion test 87
    - hypertonic-saline-infusion test 87
  - reproductive-hormone system 87
- immune system 88
- marital and sexual experience 94
- nerve-conduction velocity 74-75
- neuromotor function and exercise 73-78
  - reaction-time 73-74
    - auditory 74
    - reflex 74
    - to touch 73-74
  - tapping test 73
- nutrition 72-73
  - dietary habits 72-73
  - nutrition history 73
  - pyridoxine (vitamin-B_6) status 73
- personality and developmental characteristics 91-93
  - developmental antecedents 92-93
    - activities and attitudes questionnaire 92-93
    - parent-child-relations questionnaire 92
  - personality 91-92
    - Eysenck Personality inventory 92
    - Guilford-Zimmerman Temperament Survey 91-92
    - Imaginal Processes Inventory–Daydreaming 92
    - NEO Inventory 92
    - NEO-Rating Inventory 92
    - perceptual tests 92
    - social-desirability scale 92
- physical activity 75-78
  - muscular strength, work 75
    - coordinated exercise–cranking 75
    - physical-activity history 75
    - strength tests 75
  - oxygen uptake 75-78
    - maximal oxygen uptake 77-78
    - submaximal work loads 75-77
- pulmonary function 78-80
  - chest x-ray 80
  - gas distribution 79
  - lung and chest-wall compliance 79-80
  - maximum breathing capacity 79
  - smoking history 80
  - spirometry 78-79
- renal function 78
  - creatinine clearance 78
  - kidney, concentrating ability 78
- special senses 89
  - audiometry 89
  - eye tonography 89
  - fundus photography 89
  - visual screening 89
- stress and coping 93
  - adaptational outcomes 93
    - mood-states profile 93
    - well-being assessment 93
  - coping and defense mechanisms 93
    - coping questionnaire 93
    - coping self-interview 93
    - defense-mechanism inventory 93
  - stress 93
    - daily-events checklist 93
    - Schedule of Life-Events 93
    - stress-and-coping interview 93

Body-composition 70-72, 97-105, 171-74
- and fat 71-72, 97-99, 101-103, 135
- and sexual activity 135

Bone-alterations 25, 103-104; *App.*318
Bone-density 70-71, 99
Bone-mineral measurement 70-71, 97-99
Bonn Study 33
Breathing-capacity, maximum 56
Buerger's disease 58
Busse, Dr. E.W. 22

C

Caffeine 86, 112-13
Calcium 73, 109
Caloric intake–diet 105-10
Cameron bone-mineral analysis 70-71
Cancer 27, 38, 58, 175
Carbohydrate, in diet 109-10
  metabolism 84-86
Carbon-dioxide production 40-41
Cardiac enlargement 34
Cardiopulmonary x-ray changes 178; *App.*264
Cardiothoracic ratio 179; *App.*325
Cardiovascular disease (CVD) 3, 19, 20-21, 26-27, 30, 34, 42, 80-82, 100, 117-27, 175, 177-79, 203-205
  and intellectual performance 30
  history and physical examination 20, 26-27, 30, 34, 80-81
  Project at U. of Minn. 19, 20-21, 42
  risk factors 19, 20-21, 34, 175, 176-79
  screening 3, 20, 26-27, 80-82, 100
Cardiovascular function 20-21, 26-27, 45, 80-84, 117-27, 177-79, 203-205; *App.*272, 325
  and sexual activity 135
Carotid pulse 82
Catecholamine secretion 83, 124-25
Cell-culture senescence and *in-vitro* life-span 95-97
Cell replication 88
Cellular aging, "free-radical" theory 112
  "error" theory 112
Cellular mass, "active" 97-99
Cerebral blood-flow 30
  infarction 58
Cerebrovascular disease 34
Chicago Activity and Attitude Inventories: see "Activities and attitudes"
Cholesterol 20-21, 38, 40-41, 56, 69, 110, 165, 176-77, 208; *App.*245, 278
  as predictor of CVD 21, 34
Ciesla, Paul R. vii
Clinical "clean-up" 59-60
Clinical evaluation 63
Cognitive functions 11, 89-91, 158-62, 179-87; *App.*5, 14, 21, 25, 40, 223
Coitus 93
Cold-pressor test 20-21
Community populations, studies of 34-39
Comparison of findings across longitudinal studies 42
Compensatory change 33, 120-21, 123-24, 135, 166-67, 175, 208-209
Complaints and neuroticism 198-202; *App.*71, 79, 147
Concanavalin A 146
Concentrating ability 128
Concept-identification 91, 161-62
Coping with stress 93, 166-67
Cornell Medical Index 69-70, 164, 191, 198-200
Coronary artery disease (CAD) 3, 9, 34-36, 58, 117-27, 200-205
  and personality 200-202; *App.*71
  stress-testing for occult disease 117-18, 121-27, 204-205
Coronary events in asymptomatic subjects, prediction 204-205
Cortisone-glucose-tolerance test 85
Cranking test 75
Creatinine clearance 11, 56-58, 78, 98-99, 127-28, 170-71, 174-75, 203-204, 208; *App.*345
  nomograms 127
Criteria of selection of tests 55-57
Cross-sectional method 2, 5-9, 169-71
Cross-sectional studies: see "BLSA: cross-sectional studies"
Cross-sequential analysis 4-5, 16, 171, 182-84, 189, 194-96

D

Daily-events checklist 93
Damon, Dr. Albert 24
Dark-adaptation 25
Data-management 60-62
Data, storage and retrieval 17
Davis, Francis vii
Daydreaming 92, 162-63
  and temperament 162-63
  sex differences 163
Defense-mechanism inventory 93
Denver study 20
Dermatoglyphics of hand 70
Developmental antecedents of personality 92-93
Diabetes 3, 9, 34, 58, 128-33, 203-204
Dichotic listening 91
Diet 25, 73, 105-10, 171, 176-77; *App.*245, 372
Differential mortality 6-8
Digitalis 57, 84
Disease 1-2, 9, 19, 35, 57-60, 186, 208-209
  occult 3, 117-18, 121-22, 175
Drug metabolism 86-87, 112-14
Duke studies 22-24, 42

E

- Echocardiography 82-83
  - one-dimensional 118-19
  - two-dimensional 119
- Ejection-fraction 124
- Electrocardiogram (ECG) 20, 25, 27, 31, 34-35, 40, 56, 80-84, 117, 122, 126-27, 177-78, 204-205
  - and blood-pressure, ambulatory monitor 84, 126-27
- Ecologic study of a community 35
- Electrodermal activity 154-56
- Electroencephalogram (EEG) 19-20, 22, 74, 149, 151-54
- Electrolytes 148
- Electrophysiology 126-27
- Embedded Figure Test 92
- Endocrine system 46, 128-37
- Epidermal hydration, effect on skin-potential and vigilance 155-56
- Epinephrine, serum 83, 125
- Equivalence of information in concept-identification 161-62
- Ergometer 75
- "Error" theory of cellular aging 111-12
- Estrogens 135
- Ethanol 27, 86-87, 95, 112-14, 135, 143
  - effect on antipyrine metabolism 112-13
  - effect on "natural-killer" cell activity 143
  - effect on reaction-time and memory 86-87
  - effect on sexual activity 135
- Euglycemic clamp 86
- Exercise 11, 27, 35, 40-41, 81-82, 99-101, 121-26
- Exercise-stress test (double Master) 81-82
- Expiratory volume 31
- Eye tonography 89
- Eysenck Personality Inventory 92, 191-92

F

- Fabry's disease 58
- Family history 70
- Fat 20-21, 71-72, 97-99, 101-103, 171-73; *App.*51
- Fats in diet 109-10
- Faurot method 70
- Financing longitudinal studies 17-18
- First-visit effects 59, 170
- Flicker-fusion frequency 20
- Framingham Study 11, 34, 42
- "Free-radical" theory of cellular aging 112
- "Functional" age 26, 38, 169, 207-208
- Fundus photography 89

G

- Galvanic skin-response 149, 154-55
- Gas distribution 79
- Genetic characteristics 70
- Gerontology Research Center (GRC) vii, 3, 46, 63
- Gifted-children study 39
- Gingival disease 69
- Glomerular filtration rate: see "BLSA: cross-sectional studies, renal system;" "BLSA: longitudinal studies, renal function;" "Renal function"
- Glucose-clamp test 86, 132-33
- Glucose-insulin homeostatic system 3; *App.*387; see "BLSA: tests administered"
- Glucose-tolerance, relation to *in-vitro* cell-replication 97
- Glucose-tolerance test 35, 56, 84-86, 128-33, 203-204
  - nomograms 130, 133
- Glutamic-oxaloacetic transaminase (GOT) 110-11
- Glycolysis, anaerobic 99-101
- Gonadotropin 87, 134-35
- Granulocyte function 140-41
- Graybiel, Dr. Ashton 26
- Grip-strength 38, 41, 104
- Gsell, Dr. O.R. 30
- Growth and development 1, 5, 11-12, 19
  - studies 20
- Guilford-Zimmerman Temperament Survey 91-92, 163, 171, 189-96, 198-202; *App.*234
- Gustatory function 69, 148

H

- Haldane apparatus 72
- Handedness 70, 104
- Harvard Growth Study 20
- Hearing impairment 22
- Heart-rate 3, 41, 72, 77, 83
  - response to β-adrenergic stimulation 83
  - response to exercise 35, 41, 121-23
- Heat-production 11-13
- Height 98, 170, 172-73
- Hematocrit 69, 83
- Hemoglobin 9, 20, 38, 41, 59, 69
- Hemogram 69
- Hiroshima 36-39
- His-bundle ECG 83-84, 126
- Holtzman Inkblot Test 92
- Homosexual activity 94
- Hormone receptors 135-137
- Hormones 56, 87, 134-37
- Hyperglycemic clamp 86
- Hypertension 38-39, 59, 199-202
  - and personality 200; *App.*223
  - hypochondriasis and neuroticism 199-200
  - predictors 39
  - somatic complaints and CAD 199-202
- Hyperthyroidism 58

Hypertonic-saline-infusion test 87
Hypochondriasis 22, 198; *App*.147
Hypothalamic-neurohypophyseal-renal axis 135-37
Hypothalamic-pituitary function 3, 87
Hypothalamic-pituitary-testicular axis 134-35
Hypothalamic-posterior-pituitary function 87
Hypothenar eminence 74
Hypothyroidism 58

I

Imaginal Processes Inventory 92, 162-64
Immune response, relation to *in-vitro* cell-replication 97
Immune system 88, 97, 137-47
Immunocompetence, cellular 97
Immunoglobulins 147
Immunoproteins 147
Insulin-binding 132-33
Insulin-tolerance test, intravenous 48, 84-86, 132-33
Intelligence and blood-pressure 185-87; *App*.223
Intelligence tests 22, 39-40, 89-90, 185-87; see "Cognitive functions;" "Army Alpha test"
Interdisciplinary studies 11, 13, 17, 22-30, 95
Interpretation of data 57-60
Iodine, protein-bound, in blood 21, 25
Iowa State Univ. 39
Isoproterenol 83, 125

J

Japanese National Institute of Health 37
Japanese Radiation Effects Research Foundation 38
Johns Hopkins University Hospital 82, 83, 89

K

Kaiser Permanente Study 42
Keys, Dr. Ancel 20
Kidney 58, 78; see "Renal function"
King, Dr. John T. vii

L

Latency of involuntary and voluntary responses 154-55
L-dopa 58
Lean-body mass: see "Muscle-mass"
Learning 90-91, 158-62, 182-83; *App*.40
Life-events 23, 196-97, 202
Lipoprotein 25, 34
Liver-function 25
Longitudinal method 2, 5, 9-44, 169-71; *App*.92, 354, 362
Longitudinal studies: see "BLSA: longitudinal studies;" "Longitudinal method;" "Studies other than BLSA"
Lung and chest-wall compliance 79-80
Lung-volume 79, 114-16

M

Mackworth's Clock-Test 150-51, 154, 179-80; *App*.331
Marital and sexual history 94, 164-65
Martin, Dr. Clyde 134
Maryland State Health Dept. 80
Massachusetts General Hospital Utility Multi-Programming System 61
Masticatory function 148
Masturbation 94
Maximal breathing capacity 79
Maximal oxygen uptake 77-78
Maximum heart-rate after exercise 122-23
McMillin, P.J. vii
Medical history 56, 63
Memory 25, 87, 90-91, 158-62, 183-85; *App*.21, 25, 40, 335
Menstruation 11-13, 19
Metabolism: see "Basal metabolism;" "Drug metabolism"
Methods, uniformity 16-17
Methylphenidate 57
Metropolitan Desirable Weight Table 129
Minnesota Code 80-81
Minnesota Multiphasic Personality Inventory (MMPI) 21
Mood-states profile 93
Mortality, differential: see "Differential mortality"
Mortality studies 26-27
Multiple-gated cardiac blood-pool (MUGA) scan 83, 119-20
Muscle-mass 71, 97-98, 114, 135, 172, 174-75
Muscular function 45, 75, 157-58
Myocardial infarction 11, 80-82

N

Nagasaki 36-39
National Academy of Sciences 37
National Diabetes Data Group 85-86
National Heart Institute (NHI) vii, 34, 46
National Institute of Child Health and Human Development (NICHD) 46
National Institute of Mental Health (NIMH) Study 28-30
National Institute on Aging (NIA) 46
National Institutes of Health (NIH) vii, 29, 45, 61, 63
Natural-killer-cell activity 142-43

NEO Inventory 92, 191-92
NEO model of personality 187-94
NEO-Rating Inventory 92, 191-92
Nephrectomy 58-59
Nerve-conduction velocity 74-75
Neurology 25
Neuromotor function and exercise 73-74
Neuroticism 198-202; *App*.71, 79, 215
  and adjustment *App*.215
  and somatic complaints 198-202; *App*.71, 79
Nitro-Blue-Tetrazolium dye reduction 140
Nitrogen-washout 79, 116
Norepinephrine, serum 83, 125
Normative Aging Study 24-26, 42
Norris, Arthur H. vii
Nutrition 72-73, 105-12, 171, 176-77

O

Oakland Growth Study 20
Obesity 21, 34, 132-33, 176-77
  and longevity 21, 202-203
  and sexual activity 135
Oral-dental 25, 69, 148
Osmoreceptor sensitivity 135
Osteoarthritis of hand 104-105, 173-74; *App*.318
Osteoporosis 25, 103-104
Ouabain 84
Outcome studies: see "BLSA: outcome studies"
Oxidation, aerobic 99
Oxygen 3, 35, 40-41, 72, 75-78, 98-99, 105-10, 165, 174-75

P

Pacemaker therapy 126
Panhypopituitarism 58
Parent-child-relations questionnaire 92
Participants: see "BLSA: design and operation"
Pepsinogen in serum 25
Perceptual tests 92
Period effects 3-4, 15-16, 47, 169-71, 192-94, 208-209; *App*.92, 234
Periodontal disease 69
Peripheral-blood cells, mitogen responses of 146
Peripheral-blood lymphocyte, antibody-forming ability of 141-42
Peripheral blood, white cells in 137-40
Personality 21-22, 91-93, 166-67, 187-202; *App*.71, 79, 105, 129, 147, 163, 215, 234, 292, 304, 335
  and adjustment 197-98; *App*.129
  and cognitive performance 11
  and development 92-93
  and life-course 202; *App*.105, 129, 292
Peter, Dr. W.W. v, 3, 45-46
Phenylephrine 83
Philadelphia Geriatric Center 30
Phonocardiogram 82
Phosphorus 25
Photic "driving" 153
Photon-scan 70-71
Physical activity 34-35, 75, 99-101, 105-10, 117-18, 121-26, 177-78, 204-205; *App*.372
Physical-activity history 75
Physical examination 56, 63
"Physiologic" age: see " 'Functional' age"
Phytohemogglutinin 97, 146
Pituitary function 48
Plasma-protein electrophoresis 25
Plethysmography 84
Pokeweed Mitogen 141-42
Polymorphoneutrophile 140-41
Practice-effect 16, 116, 170, 185
Precursors of disease or death: see "BLSA: outcome studies;" "Studies other than BLSA"
Problem-solving 91, 180-82; *App*.5, 25
  and mortality 181
Propanolol 83
Proteinuria 35
Psychophysiology 148-58
Pulmonary function 11, 25, 31, 35, 40-41, 45, 78-80, 114-17, 170, 203-204; *App*.311, 387
Pulse-wave velocity 31-32
Pyridoxal phosphate, plasma (PLP) 73, 110-11
Pyridoxin 73, 110-11

Q

Quality-control 16-17, 57

R

Radiation, effect on aging 36-39
Reaction-time 38, 41, 73-74, 148-51
  and ethanol 87
Recall 90-91
Recruitment 48-51
Reflex tests 74
Regression on mean 40
Regulatory mechanisms 208
Renal function 3, 45, 57-59, 78, 127-28, 170-71, 175-76, 203-204; *App*.345, 384, 387
Reproductive-hormone system 87
Response-bias 194-96
Response-latency 75, 154-55
Response-sets 194-96
Response to stress, cardiovascular 121-26
Retinal shrinkage 25
Retroaction 161

Right-bundle branch-block 177-78; *App.*272
Risk factors 202-205; see "BLSA: outcome studies;" "Studies other than BLSA"
Rose, Dr. C.L. 24

S

Salivary-gland function 69, 148
Schedule of life events 93
Schedule of tests 55-56
Schiötz-test 89
Scientists' Cliffs, Md. vii
Sebrell, Dr. Henry vii
Sedatives 58
Sex differences in daydreaming 163
Sexual daydreams 162-63
Sexual potency 165
Shock, Dr. N.W. 47
Skin-elasticity 38
Skinfold-thickness 38, 70
Skin-potential level 154-56
Skin-potential response and vigilance 154
Smedley hand dynamometer 75, 104
Smoking 21, 27, 34, 41, 80, 85-86, 112-13, 116-17, 135, 176, 203
  and body-weight 21
  and CVD 34
  and sexual activity 135
  cigar-smoking 117
  history 26, 34, 80
  obesity, mortality 203
  pipe-smoking 117
  reversibility of pulmonary impairment 117
Social-desirability scale 92
Somatic complaints and neuroticism 198-202; *App.*79
Southern-California tests of mental ability 90
Special senses 25, 89
Speed and accuracy of movement 157
Spirometry 78-79
SRA Primary Mental Abilities tests 39
Stability 187-96, 208, 210; *App.*105, 163, 304
  definitions 187-89
  of hypochondriasis 22
  of factor-structure 192-94
  of function 208, 210
  of group-mean level 189
  of individual rank-order 189-92
  of level of sexual activity 165
  of personality 21-22, 166-67, 187-96, 208, 210; *App.*105, 163, 304
Stability-model, practical implications 196
Stability of sample, dropouts 3, 52-55
Staffing 17
Staphylococcal Protein A 146
Steering Committee 47
Stereopsis 25
Steroids 57
Strawn, Elizabeth vii
Strength tests 75
Stress and coping 93, 166-67
Stress-thallium scans 82, 204-205
Studies other than BLSA 19-42
  Basel Study 30-32
  Bonn Study 33
  CVD Project at U. of Minn. 20-21
  Duke Studies 22-24
  growth-and-development studies 5, 19-20
  incomparability of data among studies 42
  National Institute of Mental Health (NIMH) Study 28-30
  Normative Aging Study 24-26
  participants, criteria of selection 42
  Terman-Stanford study of gifted children 20
  Thousand-Aviator Study 26-28
  strengths and weaknesses 42
  studies of community populations 34-39
    ABCC Study 36-39
    Framingham Study 34-35
    Tecumseh Study 35-36
  studies of intellectual functions 39-40
  studies of physiological variables 40-41
    blood pressure 40
    exercise 40-41
  women 42
Sulfhydryl 111-12
Syphilis, serological test 69
Systolic time-intervals 82, 118

T

Tapping test 73
Taste test 70
Technetium 99m 83
Tecumseh Study 35-36, 42
Temperament 162-63, 184, 187-96; *App.*234
  and visual retention 184; *App.*335
  stability 187-96
Terman-Stanford study of gifted children 20
Testosterone 134-35
  and sexual activity 134
Tests administered: see "BLSA: tests administered"
Thematic Apperception Test 92
Thiocyanate space 71, 97-98, 172
Thomae, Dr. H. 33
Thousand-Aviator Study 26-28, 42
Thyroid function 98
Tidal volume 79
Time-perception and psychomotor function 156-57
Time-sequential analysis 4, 5, 16, 171, 182-84, 189, 194-96
Tolbutamide-response test 85-86, 130-33
  nomogram 133
Touch, reaction-time in response 73

Trait-theory 187
Tranquilizers 57-58
Transition-probability, cellular 97
Travelers' Ins. Co. 40
Treadmill-stress ECG 82
Treadmill-test 99-101
Triglycerides 25, 38, 56, 69

U

Ulnar nerve 74
United States Public Health Service (USPHS) v, vii, 34, 45, 72
Urea-blood nitrogen 25
Uremia 58
Uric acid 21, 25, 38
Urinalysis 20, 25, 38, 40, 58, 69

V

Vasodilators 57
Venous-occlusion plethysmography 84
Ventilation *App*.311
Ventilatory response to exercise 35
Ventricular hypertrophy 120-21
Verbal ability 39; see "Vocabulary (WAIS)"
Verbal learning 90, 158-61
Verzar, Dr. F. 30
Veterans' Administration Study: see "Normative Aging Study" under "Studies other than BLSA"
Vibratory perception 38
Vigilance 148-51, 154, 179-80; *App*.331
Visual memory 183-84; *App*.14, 21, 40
Visual retention and temperament 184-85; *App*.335
Visual screening 22, 25, 32, 38, 89
Vital capacity 32, 34, 38, 41, 56, 78-79
Vitamin-A 109
Vitamin-B_6 73, 110-11
Vitamin-C 109
Vitamin allowances recommended by National Research Council 109
Vocabulary (WAIS) 90, 160, 183-84, 191

W

Watt, Dr. James vii
Waxter, Dr. T.J.S. vii
Wechsler Adult Intelligence Scale (WAIS) 33, 90, 160, 183-84, 191
Weight 169-70, 173; *App*.51, 372
Well-being 93; *App*.129
Wernick, Paula vii
White cells in peripheral blood 137-40
Wilmer Eye Institute 89
Women participants: see "BLSA: design and operation;" "Studies other than BLSA"
Wyman, Mr. and Mrs. William E. 48

X

X-ray, chest 20, 25, 38, 40, 80, 178-79; *App*.264
  dental 69
  hand 25, 70, 103-105

Y

"Your Activities and Attitudes:" see "Activities and attitudes"